UNIVERSITÉ LOUIS PASTEUR STRASBOURG

Laboratoire des Sciences
de l'Image, de l'Informatique
et de la Télédétection

UMR-CNRS-ULP 7005

Équipe Modèles Image Vision

Thèse presentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur
Strasbourg I

Discipline : Informatique

par Erchan Aptoula

Analyse d'images couleur par morphologie mathématique. Application à la description, l'annotation et la recherche d'images

Soutenue publiquement le 10 Juillet 2008

Membres du jury

Directeur de Thèse :	Christian RONSE, Professeur Université Louis Pasteur, Strasbourg
Rapporteur Interne :	Mohamed TAJINE, Professeur Université Louis Pasteur, Strasbourg
Rapporteurs Externes :	Christine FERNANDEZ-MALOIGNE, Professeur Université de Poitiers, Poitiers
	Jean SERRA, Directeur de Recherche émérite ESIEE, Noisy-le-Grand
Examinateurs :	Sébastien LEFÈVRE, Maître de Conférences Université Louis Pasteur, Strasbourg
	Bülent SANKUR, Professeur Université de Boğaziçi, İstanbul

To my family

Acknowledgements

It has been almost three years since the day I started this thesis, and I must admit that it has been quite an adventure, with several exciting moments, including losing and regaining funding. Even though this manuscript represents my personal work, it would either have been very different or not possible at all, without the contribution of the following people, whom I feel obliged to thank here.

Christian Ronse, my thesis advisor, who accepted to supervise my work, despite my not particularly promising M.Sc. exam results, as well as the distance of the subject to his main research interests.

Sébastien Lefèvre, who made sure I stayed on track, whenever I misused the freedom of organising my efforts, and ensured excellent working conditions.

Christine Fernandez-Maloigne, Jean Serra and Mohamed Tajine, who have undertaken the cumbersome task of judging this manuscript.

Bülent Sankur for accepting to be member of the jury, despite his busy work schedule.

In addition to these people that have had a major influence on my academic work, I would also like to thank those that have had a rather indirect, yet considerable effect on my stay in France.

Alain Ketterlin, for his support, especially during my early days in Strasbourg.

My friends and collegues in, but not limited to LSIIT, Ahmed Ahmed, Chadi Albitar, Tuna Alkan, Abdelkrim Belhaoua, Ozan Erdem Bişgin, Bessem Bouraoui, Süha Demir Can, Begüm Demir, Sébastien Derivaux, Özgür Develioğlu, Sezen Duygu, Volkan Erol, Bülent Hakkıoğlu, Redouane Khemmar, Zafer Kızıltoprak, Daeyun Kim, Haralambos Michailidis, Ahmed Mümtaz, Benoît Naegel, Dimitra Papadimitriou, Nicolas Passat, Benjamin Perret, George Sfikas, Florent Sollier, Bahar Şahin, Selçuk Takır, Selman Tayyar, Gülen Türkeri, Victoria and Jonathan Weber, who have made my stay in Alsace much more enjoyable through their company.

Finally, I wish to specially thank my family for their unconditional love and support, despite being far away for so long.

Abstract

This thesis concentrates mainly on the extension of mathematical morphology to colour images. The resulting morphological operators are then applied to the problem of general purpose content-based colour image retrieval.

First the choice of a colour space is discussed and polar spaces of type luminance-saturation-hue are endorsed on the grounds of their intuitiveness, as far as the human visual perception is concerned. The rest of the first part focuses on the sole requirement for the extension of mathematical morphology to multivariate data: a vector ordering. Specifically, due to its multiple positive properties, we concentrate on the lexicographical approach. We further introduce several variations to it, with the purpose of rendering possible a finer tuning of the processing balance among image channels, and between chromatic and achromatic information in the particular case of colour images. Their practical interest is illustrated through a comparative study, that is carried out on the properties of several vector orderings. Moreover, considering the periodic nature of hue values, that does not lend itself to scalar ordering, an image specific approach based on multiple reference hues is proposed. This part is concluded with the extension of the hit-or-miss transform, a generic template matching tool, to colour images, as an example based on its findings.

The second part is dedicated to the application of colour morphology to the problem of content-based image retrieval. To this end, novel morphology based colour and texture descriptors, using the principles developed in the first part, are presented and compared against state-of-the-art approaches with multiple image databases. In particular, multiscale histogram variations are studied as global colour descriptors, whereas the information extracted by granulometry and morphological covariance is combined with the purpose of texture characterisation. Finally a semi-supervised, keyword-based image annotation and retrieval architecture is developed, based on the aforementioned operators, leading to a flexible solution in terms of adaptability to various image collection heterogeneity levels, providing feedback support and capable of semantic queries.

Keywords: colour image analysis, multivariate mathematical morphology, vector orderings, image description, feature extraction, content-based image retrieval.

Résumé

Cette thèse porte principalement sur l'extension de la morphologie mathématique aux images couleur. Les opérateurs morphologiques qui en dérivent sont ensuite appliqués au problème de la recherche d'images par le contenu.

L'étude de ce problème débute avec le choix d'un espace de couleurs "pertinent", et il a été décidé d'utiliser les espaces de couleurs s'appuyant sur un triplet luminance-saturation-teinte, plébiscités pour leur intuitivité et leur bonne représentation du système de la vision humaine. Suite au choix d'espace de couleurs, nos efforts se sont portés sur le seul prérequis pour permettre l'extension de la morphologie mathématique aux données multivariées : un ordre et des opérateurs d'extremums pour des données vectorielles. Plus précisément, nous nous sommes intéressés à l'approche lexicographique, du fait de ses propriétés théoriques intéressantes et de sa capacité d'adaptation à différents contextes applicatifs. Plusieurs solutions ont été proposées pour s'affranchir de son principal inconvénient qui consiste en la prioritisation extrême de la première bande spectrale ou couleur. En outre, une attention particulière a été portée au traitement de la teinte, susceptible de poser des problèmes si elle est ordonnée comme une valeur scalaire alors que sa nature est périodique. Par conséquent nous présentons une approche utilisant plusieurs teintes de référence. Cette partie se termine par l'extension aux images multivariées de la transformée en tout-ou-rien, découlant des résultats obtenus précédemment.

La deuxième partie de ce travail concerne l'application de la morphologie couleur au problème de la recherche d'images par le contenu. Dans ce but, nous avons introduit plusieurs descripteurs globaux de couleur et de texture. Plus précisément, nous nous sommes intéressés au concept d'histogramme multi-échelles et avons élaboré deux méthodes permettant leur intégration avec les opérateurs morphologiques. En outre, des granulométries spécifiques aux images couleur ont aussi été étudiées. Pour la description de texture, nous avons combiné les deux principaux outils morphologiques de caractérisation de texture, la granulométrie et la covariance. Finalement, nous avons proposé l'architecture d'un système pour la description, l'indexation et la recherche par le contenu à base de mots-clés, qui fournit une solution flexible en termes d'adaptabilité aux différents niveaux d'hétérogénéité de collections d'images.

Mot-clés: analyse d'images couleur, morphologie mathématique multivariée, ordres vectoriels, description d'images, extraction de caractéristiques, recherche d'images par le contenu.

Özet

Bu tez çalışması, matematiksel biçimbilimin renkli görüntülere uzantısını konu almaktadır. Elde edilen biçimbilimsel işleçler daha sonra genel amaçlı içerik tabanlı görüntü geri alımına uygulanmıştır.

Öncelikle renk uzayı seçimi konusu ele alınmış, ve aydınlık-doygunluk-renközü tabanlı uzaylar üzerine, insan görüsünün ilkelerine olan yakınlıkları nedeniyle yoğunlaşılmıştır. İlk kısmın kalanı, matematiksel biçimbilimin renkli görüntülere olan uzantısı için gerekli tek oğeye odaklanmaktadır: yöneysel bir sıralama. Olumlu kuramsal özellikleri nedeniyle bu konuda sözlüksel sıralama seçilmiştir. Görüntü kanallarının, ve özellikle renkli ile renksiz verinin, daha dengeli işlenmesini sağlamak adına, sözlüksel sıralamaya ayrıca birkaç değişiklik önerilmiştir. Uygulamadaki başarımları, birçok yöneysel sıralamanın dahil olduğu karşılaştırmalı bir inceleme ile gösterilmiştir. Bundan başka, renközü değerlerinin devirli doğası nedeniyle, sayıl sıralamaya uygunsuzlukları göz önünde bulundurularak, görüntüye özel ve birden fazla taban renközü değerine dayalı bir yaklaşım sunulmuştur. Bu kısım, sonuçlarının bir örneği olarak, ıskala-yakala dönüşümünün renkli görüntüler için tanımı ile tamamlanmıştır.

İkinci kısım, renkli biçimbilimin içerik tabanlı görüntü geri alımına uygulanmasına ayrılmıştır. Bu amaçla, ilk kısmın sonuçlarına dayanan, yeni ve biçimbilim tabanlı renk ve doku tanımlayıcıları sunulmuş ve bu alanda tanınan yaklaşımlar ile karşılaştırılmışlardır. Özellikle çok ölçekli dağılım eğrileri küresel renk tanımlayıcıları olarak, ve tanecikölçüm ile biçimbilimsel değişinti birleşimi doku tanımlayıcısı olarak incelenmiştir. Son olarak, tasarlanmış işleçlere dayalı, yarı-eğitimli, anahtar sözcük tabanlı ve farklı türdeşlik düzeyindeki görüntü veritabanlarına uyum sağlayabilen esnek bir görüntü geri alım mimarisi geliştirilmiştir.

Anahtar sözcükler: renkli görüntü işleme, yöneysel matematiksel biçimbilim, yöneysel sıralamalar, görüntü tanımı, öznitelik çıkarımı, içerik tabanlı görüntü geri alımı.

Résumé étendu

1. Problématique

Cette thèse porte principalement sur l'extension de la morphologie mathématique aux images couleur. Les opérateurs morphologiques qui en dérivent sont ensuite appliqués au problème de la recherche d'images par le contenu.

L'intérêt des opérateurs morphologiques univariés dans de nombreux domaines, couplé au besoin croissant de disposer de traitements adaptés aux images multivaluées de plus en plus nombreuses, représentent les principales motivations de ce travail doctoral focalisé sur l'extension de la théorie de la morphologie mathématique aux données multivariées. Le faible nombre de prérequis pour permettre cette extension (seuls un ordre et des opérateurs d'extremums pour des données vectorielles sont nécessaires) a donné lieu à un nombre important de propositions dans la littérature, sans toutefois aboutir à une solution satisfaisante. La première partie de cette thèse porte donc sur ce problème et débute par une discussion relative au choix d'un espace de couleurs pertinent pour les traitements morphologiques. Nous nous intéressons en outre au problème principal de l'analyse morphologique d'images couleur, l'ordonnancement des vecteurs de couleur nécessaire pour obtenir une structure de treillis complets sur l'ensemble des pixels multivariés.

La deuxième partie de ce travail concerne l'application de la morphologie couleur au problème de la recherche d'images par le contenu. Ce domaine de recherche récent (les principales méthodes rapides et efficaces de recherche par le contenu des données visuelles datent du début des années 90) est très actif, notamment du fait de l'explosion de la quantité d'images produites et gérées par les utilisateurs (fonds personnel ou d'entreprise, internet). Cependant, bien que la morphologie mathématique ait démontré son potentiel pour l'extraction de caractéristiques et la description du contenu dans plusieurs domaines, elle reste largement sous-exploitée pour traiter le problème de la recherche par le contenu, à l'exception de cas très particuliers (classification des cellules hématologiques par exemple). Dans cette partie de la thèse, les contributions portent principalement sur le développement de descripteurs morphologiques du contenu des images en s'appuyant sur les résultats obtenus dans la première partie.

2. Espaces de couleurs

Bien que les propriétés physiques de la lumière, qui sont à l'origine de la perception de la couleur soient étudiées depuis plusieurs siècles, les mécanismes internes, et en particulier les étapes avancées de la vision humaine ne sont pas entièrement connus. Par conséquent, le premier obstacle à l'extension de la morphologie mathématique aux images couleur, consiste en le choix d'une représentation effective de cette information. Nous avons débuté notre étude par un recensement des principaux espaces de couleurs, chacun possédant des propriétés diverses, rendant ainsi certains plus adéquats que d'autres pour le traitement d'images couleur. Dans ce but, suite à un rappel des principes de la perception humaine de la couleur, nous avons étudié les espaces suivants : RVB (RGB)[1], espaces perceptuellement uniformes (L*a*b*, L*u*v*), espaces spécifiques à certaines applications (CMJN (CMYK), Y'UV) et espaces polaires (TSV (HSV), TLS (HLS), L*C*T* (L*C*H*)).

Les espaces perceptuellement uniformes ainsi que polaires ont été particulièrement étudiés. En effet, les premiers offrent l'avantage d'un calcul de distance entre couleurs conforme à la perception humaine, tout en garantissant l'indépendance aux conditions d'acquisition, à un coût de conversion cependant relativement élevé. Les espaces polaires, modélisent la perception humaine de la couleur du point de vue psycho-visuel, mais les propriétés mentionées ci-dessus en sont absentes. Les défauts des versions cylindriques de TSV et de TLS ont également été soulignés, et l'espace LST (LSH) qui s'en affranchit a été étudié.

Comme le choix optimal d'un espace de couleurs dépend fortement des besoins de l'application considerée, suite à la présentation des propriétés offertes par les différents espaces, nous avons évalué l'adéquation de ces espaces à notre cas d'étude, en d'autres termes, le traitement d'images couleur par la morphologie mathématique.

3. Morphologie mathématique multivariée

La morphologie mathématique est un cadre particulièrement riche pour l'analyse d'images, principalement développé pour les images binaires et à niveaux de gris. Sa popularité dans la communauté du traitement d'image est due principalement à une base mathématique rigoureuse et une capacité implicite à exploiter les relations spatiales entre les pixels. Cependant, son extension aux images couleur, ou plus généralement aux données multivariées, n'est pas triviale.

Plus précisément, dans le cadre morphologique les images sont représentées par des fonctions de $\mathcal{E}$, l'espace des coordonnées (souvent $\mathbb{R}^d$, l'espace Euclidien d-dimensionel ou $\mathbb{Z}^d$ l'espace d-dimensionel discret), dans $\mathcal{T}$ l'ensemble des valeurs des pixels, un treillis complet. En d'autres termes, $\mathcal{T}$ est un ensemble non vide associé à un ordre partiel, où tout sous-ensemble non vide $\mathcal{P}$ de $\mathcal{T}$, a un supremum (plus petit majorant) et un infimum (plus grand minorant). En prenant l'ensemble d'images $\mathcal{F} = \mathcal{T}^{\mathcal{E}}$, la même structure de treillis est imposée à $\mathcal{F}$:

$$f, g \in \mathcal{F}, \;\; f \leq g \Leftrightarrow \forall\, x \in \mathcal{E}, \;\; f(x) \leq g(x)$$

Une fois cette structure verifiée, il est facile de définir les opérateurs de base, l'érosion et la dilatation d'une image f avec un élément structurant b, respectivement par :

$$\begin{aligned} \varepsilon_b(f)(\mathbf{x}) &= \inf_{v\,\mathbf{s} \in b} \{f(\mathbf{x}+\mathbf{s})\} \\ \delta_b(f)(\mathbf{x}) &= \sup_{v\,\mathbf{s} \in b} \{f(\mathbf{x}-\mathbf{s})\} \end{aligned}$$

où les extremums sont calculés à partir de l'ordre en question. Dans le cas d'images binaires ($f : \mathbb{R}^d \to \{0, 1\}$), la relation d'ordre est représentée par la relation d'inclusion, alors que dans le cas d'images à niveaux de gris ($f : \mathbb{R}^d \to \overline{\mathbb{R}}$), c'est l'ordre scalaire dans $\overline{\mathbb{R}}$ qui est

[1]Abbreviation anglaise

utilisé. Quant aux images multivariées ($f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}^n$, $n > 1$) par contre, il n'y a pas une methode universelle d'ordonnancement. Par conséquent, la difficulté majeure de l'extension des opérateurs morphologiques aux images couleur est l'ordonnancement des couleurs au moyen d'un ordre vectoriel.

Le problème d'ordonnancement de données vectorielles a été posé depuis de nombreuses années, et plusieurs solutions ont été proposées, et ont pour la plupart été utilisées pour définir de nouveaux opérateurs morphologiques multivariés ; cependant aucune n'a été unanimement acceptée. Les approches d'ordonnancement vectoriel dans le contexte de la morphologie couleur peuvent etre regroupées selon la classification suivante.

L'ensemble des ordres partiels inclut l'ordre marginal où chaque composante d'un vecteur est ordonnée indépendamment des autres. Par conséquent toute information corrélationelle est ignorée, alors que la répétition de l'application d'opérateurs pour chaque bande de l'image entraîne un coût de traitement supplémentaire. En outre, du fait de sa nature partielle, il peut y avoir des vecteurs non comparables marginalement, ce qui conduit à des extremums qui n'existent pas toujours dans l'image d'entrée, un phénomène connu sous le nom de "fausses couleurs". Par contre, un tel ordre permet l'utilisation directe des opérateurs morphologiques définis pours les images à niveaux de gris, sans besoin d'adaptation aux images couleur.

L'ensemble des pré-ordres totaux, évite le problème de fausses couleurs en préservant les vecteurs d'entrée, grâce à la propriété de totalité. Les pré-ordres se basent souvent sur la scalarisation des vecteurs, c'est-à-dire la transformation de vecteurs en scalaires, puis l'ordonnancement selon l'ordre scalaire des réels. Ainsi ils offrent la possibilité de manipuler les dimensions des vecteurs, ou les bandes des images couleur, de façon symétrique. Néanmoins, étant des pré-ordres, ils ne satisfont pas la propriété d'anti-symétrie, et il peut donc y avoir des vecteurs distincts qui sont considérés equivalents, ce qui reduit les propriétés théoriques des opérateurs qui en résultent.

L'ensemble des ordres totaux, préserve non seulement comme précédemment les vecteurs d'entrée grâce à sa totalité, mais évite également le problème de multiples extremums, puisqu'il vérifie la propriété d'anti-symétrie. Le représentant le plus connu de cette classe est l'ordre lexicographique qui consiste à ordonner des vecteurs dimension par dimension, en commençant par la première, et en ne traitant la suivante qu'en cas d'égalité. Cet ordre a été utilisé dans plusieurs publications afin de définir des opérateurs morphologiques, en association avec des espaces de couleurs qui offrent une priorité naturelle entre leur dimensions. La facilité de configuration, (obtenu en choisissant l'ordre de dimensions à comparer) est un avantage supplémentaire. Cependant son principal inconvénient est la prioritisation extrême de la première dimension, à un tel niveau que les dimensions suivantes ne sont que guère exploitées.

Suite à ce panorama des approches d'ordonnancement vectoriel, nous avons réalisé une série d'expériences pour mieux comprendre leurs différences pratiques, et illustrer au moyen d'exemples leurs propriétés théoriques. Plus précisément, plusieurs ordres représentatifs de chaque classe ont été comparés à l'aide d'applications en réduction de bruit et en classification de textures, en utilisant différents espaces de couleurs. En résumé, il a été constaté que l'ordre marginal est le plus efficace pour réduire le bruit non corrélé, alors que les performances de différents ordres sont beaucoup plus proches dans le cas de bruit correlé. L'ordre lexicographique a d'ailleurs montré une performance souvent supérieure, en réduction de bruit ainsi qu'en classification, quand il est couplé avec un espace polaire, conformément à la littérature où cette combinaison est relativement populaire. Après avoir étudié les espaces de couleurs, ainsi que les approches d'ordonnancement, nous avons conlu cette partie avec le choix d'un couple (ordre vectoriel, espace de couleurs).

En ce qui concerne le choix d'un ordre vectoriel, nous avons constaté que chaque ordre possède des propriétés diverses, les rendant plus appropriés pour certaines tâches. Par exemple, les pré-ordres totaux peuvent fournir un traitement symétrique des bandes d'une image, alors que l'ordre lexicographique est plus adapté au cas où une priorité entre les bandes existe. Par conséquent, aucune ordre ne peut être consideré comme approprié pour tout type de traitement couleur, puisque les besoins pratiques varient énormément. Donc, à l'exception des cas où les besoins pratiques sont bien définis et connus a priori, la capacité d'adaptation de l'approche d'ordonnancement est necessaire.

En outre, les approches qui n'assurent pas certaines propriétés fondamentales telles que l'anti-symétrie (qui ne pouvant donc même pas être qualifiées comme des ordres au sens algébrique du terme, et qui ne garantissent pas l'unicité des extremums) ont été écartées pour la suite. En outre, même si certaines applications comme la réduction de bruit ne l'imposent pas, la préservation des couleurs est un point supplémentaire à prendre en compte, incitant ainsi à nous orienter vers les ordres totaux, et en particulier l'ordre lexicographique, que nous avons comme base à notre cadre morphologique couleur.

Quant au choix d'un espace de couleurs, nous avons observé qu'un grand nombre d'entre eux a été déjà utilisé pour définir des opérateurs morphologiques couleur. Conformément à l'étude très complète de Serra, Hanbury et Angulo [9, 95, 243, 244] il a été constaté que les espaces polaires, du fait de leur représentation intuitive des couleurs, en ce qui concerne l'observateur humain, constituent un choix populaire dans ce domaine. L'uniformité perceptuelle par contre, fourni par exemple par l'espace $L^*C^*T^*$, a été considerée comme une propriété non cruciale, de même que l'indépendance aux conditions d'acquisition, étant donné que celles-ci ne sont que rarement connues. En outre, puisque nous avons souligné l'intérêt du principe de flexibilité de point de vue de l'approche d'ordonnancement, nous attachons une importance particulière aux espaces de couleurs compatibles avec ce principe. Pour toutes ces raisons, nous avons décidé de nous focaliser sur les espaces de couleurs polaires, et en particulier l'espace LST.

4. Un ordre lexicographique dans l'espace LST

Ayant choisi un couple d'un ordre vectoriel et d'un espace de couleurs, nos efforts se sont ensuite portés sur l'étude plus approfondie de cette configuration. Bien que des opérateurs morphologiques théoriquement corrects puissent être définis de cette façon, leur intérêt pratique reste relative limité, deux points demandent une attention particulière restant encore à traiter. Il s'agit d'abord du traitement de la teinte, qui pose des problèmes puisqu'il faut prendre en compte sa nature périodique. En outre le niveau de priorité attribué à la première dimension vectorielle pendant l'ordonnacement lexicographique, est souvent si élevé qu'il conduit à une sous-exploitation très forte des autres bandes disponibles. Nous nous sommes donc focalisés sur chacun de ses problèmes indépendamment et y avons apportés des solutions appropriées.

4.1 Traitement de la teinte

La teinte est susceptible de poser des problèmes en termes de traitement morphologique, étant donné qu'il est difficile d'imposer une structure de treillis sur l'ensemble de teintes, qui sont des valeurs angulaires définies sur $[0, 2\pi]$. Ainsi, la teinte d'origine $h = 0$ qui correspond au rouge peut conduire à l'apparition de discontinuités dans les images, au cas où l'ordre scalaire est utilisé. C'est pourquoi, depuis quelques années, et suite aux travaux de Peters [207] et d'Hanbury et Serra [103], une approche alternative a été préférée. Elle consiste à ordonner les teintes au moyen d'un pré-ordre basé sur les distances à une teinte de référence. Plus précisément, une teinte est choisie comme référence, automatiquement où par un expert, cette référence étant censé représenter la teinte dominante de l'image. Ensuite toute valeur sur le cercle de teintes est ordonnée selon sa distance angulaire à cette référence, et en particulier, une teinte est considerée d'autant plus grande que sa distance à la référence est petite.

L'utilisation d'une teinte de référence dynamique est certainement une approche intéressante pour imposer une structure de treillis sur le cercle de teintes. Par contre, elle présente également certains inconvénients. Plus précisément, en choisissant une teinte d'origine arbitraire, les opérateurs morphologiques de base (erosion et dilatation) vont fournir une teinte soit plus proche, soit plus éloignée de cette teinte de référence. En prenant par exemple le vert comme référence, on aura vert superieur à jaune et cyan. Cependant, l'ordonnancement d'autres teintes comme rouge et magenta sera défini de façon implicite selon leur position sur le cercle de teintes et selon leur distance au vert. Cela peut conduire à des résultats peu intuitifs pour les images où plusieurs teintes dominantes existent, ce qui est souvent le cas en pratique. C'est pourquoi, nous proposons d'utiliser une variante de cette approche, en considerant plusieurs teintes de référence représentatives du contenu de l'image. Etant donné l'ensemble des teintes dominantes, les teintes sont ordonnées selon leur distance par rapport à ces références, et en particulier, la valeur qui est la plus proche d'une teinte dominante de l'image est considerée comme la plus élevée. Nous proposons également une approche permettant à obtenir de façon automatique les teintes de référence, qui peuvent être ponderées pour contrôller l'effet des références sur l'ordonnancement selon leur présence dans l'image.

4.2 Variantes de l'ordre lexicographique

Afin de mieux contrôler la transition de comparaison parmi les dimensions vectorielles pendant l'utilisation de l'ordre lexicographique, trois approches ont été élaborées. Nous nous sommes d'abord intéressés à l'ordre lexicographique α-modulus, qui constitue une solution efficace à ce problème. Elle consiste à quantifier la dimension à l'aide d'une division par a suivie d'un arrondi à l'entier supérieur. Ainsi les égalités au sein de la première bande sont plus fréquentes, permettant une transition à la deuxième bande. Néanmoins, cette approche réalise une sous-quantification homogène de la première bande, or cette hypothèse est relativement arbitraire. Dans le cas des images couleur, et en particulier des espaces polaires, il existe des relations couleur plus complexes entre les bandes. Plus précisément, du fait de la forme bi-conique de l'espace de couleurs, la saturation n'a une l'importance que pour des valeurs non-extrêmes de luminance, alors que la teinte n'a de sens que pour des valeurs de saturation "suffisament" élevées. Afin de modéliser ces relations dans le contexte d'un ordonnancement de couleurs dans l'espace LST, nous avons proposé une généralisation de l'ordre lexicographique α-modulus où des distributions de priorité arbitraires peuvent être utilisées. De plus, des modèles de pondération pour les couples de luminance-saturation et saturation-teinte ont été élaborés, et utilisé judicieusement dans le nouvel ordre lexicographique tout en préservant ses propriétés théoriques.

Une deuxième variante proposée de l'ordre lexicographique, se base également sur la notion de sous-quantification mais cette fois dans le domaine spatial et non spectral. Plus précisément, même si des approches sophistiquées de sous-quantification spectrale sont utilisées, les discontinuités spatiales dans les images sont inévitables et a priori inconnues. Nous avons étudié dans ce but, l'utilisation d'une image auxiliaire à niveaux de gris appelée "marqueur". Étant donné un couple de pixels couleur dont on compare la dimension i, on peut utiliser les pixels de mêmes coordonnées dans l'image marqueur, contenant plus de régions plates du point de vue topologique, augmentant ainsi la fréquence de transition de la cascade lexicographique aux bandes ultérieures. Les régions plates peuvent être obtenues par plusieurs opérateurs morphologiques à niveaux de gris, par exemple les nivellements.

La troisième approche proposée pour affiner les effets de l'ordre lexicographique est plutôt peu conventionelle. Elle ne consiste pas en la définition d'un ordre ensuite utilisé pour calculer des extremums vectoriels nécessaires à l'érosion et la dilatation, mais porte directement le calcul final d'extremums vectoriels, étant donné un ensemble de vecteurs. Cette approche présente l'avantage d'exploiter la distribution spectrale des vecteurs pendant le calcul d'extremums, pour mieux contrôler la transition parmi les bandes. Par contre, en l'absence d'ordre du point de vue algébrique, cette méthode conduit à des opérateurs pseudo-morphologiques.

Plus précisément, elle consiste à utiliser le principe de α-troncature, longtemps utilisé en filtrage d'image. Étant donné un ensemble de vecteurs, on ordonne les vecteurs selon leur première dimension, en utilisant l'ordre scalaire. Ensuite on garde un pourcentage (défini par une variable α) des plus grands vecteurs, et on répète l'opération pour les dimensions ultérieures sauf la dernière. Ainsi, la procédure se termine soit avec un seul vecteur qui est considéré comme le supremum de l'ensemble ou avec plusieurs, et dans ce cas l'ordre scalaire selon la dernière dimension fournit le supremum. En outre, une approche basé sur l'écart-type de valeurs au sein de chaque bande est proposée pour calculer la variable α de façon automatique. De plus, un algorithme est presenté pour fournir un ordre au sens algébrique à partir de cette approche. Il consiste à calculer des extremums itérativement étant donné un espace de vecteurs. Une fois le premier extremum calculé, celui-ci est supprimé de l'ensemble de calcul, et l'étape est répétée.

Ces trois approches ont été évaluées au moyen des applications en réduction de bruit et en classification de texture, et l'approche d'α-troncature a conduit à des resultats surprenant malgré ses limites théoriques, tandis que l'approche de sous-quantification généralisée a fourni une performance supérieure à l'ordre lexicographique α-modulus.

4.3 Transformée en tout-ou-rien vectorielle

Pour clore cette première partie de la thèse, nous nous sommes finalement intéressés à la transformée en tout-ou-rien, un outil morphologique de reconnaissance de motifs de base. Elle consiste à chercher et détecter un motif dans une image, étant donné un masque. En se basant sur les résultats précédents, nous avons fourni une définition pour une transformée en tout-ou-rien multivariée, qui peut détecter des motifs multivariés. Plus précisément, il suffit de remplacer dans sa définition à niveaux de gris les opérateurs scalaires par leur homologues vectoriels, en imposant en outre un ordre invariant par translation tel que l'ordre lexicographique. Le paramètre supplémentaire de l'ordre vectoriel permet d'affiner la capacité de détection de l'opérateur obtenu. En outre, deux techniques classiques utilisées pour rendre la transformée en tout-ou-rien plus flexible, les éléments structurants synthétiques et les filtres de rang, ont été adaptés au cas vectoriel. Cependant, l'intérêt pratique de cet outil pour les images couleur, en ce qui concerne la détection d'objets couleur dans un cadre général, reste encore très limité, du fait de son manque d'invariance aux changements d'échelle et aux rotations.

Pour la suite de cette thèse, nous avons choisi d'utiliser l'approche proposée de traitement de teinte ainsi que l'ordre lexicographique basé sur la sous-quantification pour réaliser tout traitement morphologique couleur.

5. Recherche d'images par le contenu

L'explosion de la masse des informations visuelles, résultat de l'augmentation constante des capacités des dispositifs d'acquisition et de la décroissance des coûts des moyens de traitement et de stockage, a conduit à la formation d'entrepôts d'information particulièrement riches. Des outils robustes et automatiques pour leur gestion, recherche et annotation sont donc nécessaires. Néanmoins, les systèmes de gestion de bases de données qui ont été developpés pour le problème des entrepôts d'informations textuellues, ne peuvent pas être utilisés de façon efficace dans ce contexte, puisqu'ils ont été principalement conçus pour des données alpha-numériques, tandis que les images, du fait de leur nature visuelle, nécessitent de préférence un accès par leur contenu.

Dans ce but, depuis les années 90, plusieurs systèmes de recherche d'images par le contenu ont été élaborés. Dans un tel système, les images sont décrites par des vecteurs de caractéristiques, représentés par des vecteurs de réels, calculés à partir de descripteurs. Cette description numérique et compacte vise à capturer le contenu de l'image, en stockant des informations sur la distribution des éléments visuels de base, comme la couleur, la texture et la forme. Ainsi on peut manipuler les images, et faire des recherches. Néanmoins, le défi majeur de ce domaine continue à être la decription efficace du contenu sémantique des images. Plus précisément, après les études initiales de faisabilité, le secteur industriel ainsi que la communauté scientifique se sont focalisés sur les besoins des utilisateurs. Or ces utilisateurs, en majorité, n'effectuent presque jamais des requêtes qui consistent en des critères d'éléments visuels de base, par exemple "toutes les images qui contiennent 70% de rouge et des formes circulaires de texture X". Au contraire, les utilisateurs veulent être capables de rechercher des notions sémantiques (couché du soleil), des objets (bâtiment), voire même des concepts abstraits (paix). Cependant, les besoins des utilisateurs ne sont pour l'instant pas satisfaits par les système de recherche par le contenu. Cette distance entre les éléments visuels de base et les notions sémantique de haut niveau s'appelle "fossé semantique" ou "semantic gap", et constitue le défi majeur du domaine.

Après avoir dressé un panorama des travaux dans le domaine de la recherche d'images par le contenu nous nous sommes focalisés sur l'application de la morphologie mathématique à ce problème.

6. Morphologie mathématique et la description d'images par le contenu

Bien que l'intérêt de la morphologie mathématique ne soit plus à démontrer depuis longtemps pour de nombreux problèmes de traitement d'images, son application dans le domaine de la recherche d'images par le contenu est récente et relativement limitée. Les quelques résultats obtenus à ce jour permettent néanmoins d'observer que les opérateurs morphologiques sont effectivement pertinents dans ce domaine, en particulier du fait de leur capacité à exploiter les relations spatiales entre les pixels.

Dans le prolongement des travaux menés lors de la première partie, nous nous sommes intéressés à la description d'éléments visuels de base au moyen d'outils morphologiques, et avons étudié en particulier la couleur et la texture. La forme n'a pas fait l'objet de travaux particuliers, puisqu'elle constitue une notion binaire et nécessite une segmentation automatique. En ce qui concerne la couleur trois descripteurs globaux ont été proposés.

Nous avons d'abord introduit les granuloméries spécifiques aux données couleur. Elles décrivent indépendamment les distributions d'intensité ainsi et de taille de chaque couleur, ajoutant l'information spatiale à la description qui en résulte. En outre, nous nous sommes intéressés au concept d'histogramme multi-échelles et avons élaboré deux méthodes permettant leur intégration avec les opérateurs morphologiques. Nous avons étudié l'efficacité des histogrammes multi-échelles basés sur des espaces d'échelle obtenus par des nivellements couleur (un opérateur morphologique puissant fournissant des régions plates) d'une part, et par le processus de segmentation par ligne de partage des eaux d'autre part.

Quant à la description de la texture qui a été abordée par la suite, nous avons combiné, au moyen de plusieurs variables d'éléments structurants, les deux principaux outils morphologiques de caractérisation de texture, la granulométrie et la covariance, afin de mieux exploiter l'information complémentaire fournie par chacun d'eux.

Les descripteurs ont été comparés à des alternatives connues, en termes de performance de description, ainsi que de robustesse au bruit. Les résultats obtenus ont montré que les opérateurs morphologiques peuvent fournir une performance équivalente, avec néanmoins une complexité de calcul plus importante.

7. Un système de recherche d'images par le contenu : MIMAR

Finalement, en nous appuyant sur les méthodes morphologiques de description du contenu développés précédemment, nous avons proposé l'architecture d'un système pour la description, l'indexation et la recherche par le contenu à base de mots-clés. La conception du système a été réalisée selon des restrictions imposées dans le cadre d'un contrat avec Oséo, avec en particulier une contrainte forte quant à l'adaptabilité du système à la variabilité des bases d'images, tout en étant capable de réaliser des requêtes de nature sémantique.

Dans ce but, étant données les performances des descripteurs élaborés, il a été décidé d'apporter la connaissance sémantique de façon exogène, en d'autres termes par un expert. Celui-ci est censé connaître le domaine relatif à la base d'images, et fournit ses connaissance sous forme d'images exemples et contre-exemples, pour une notion donnée. Cette notion, représenté par un mot-clé, est ensuite associée avec une description vectorielle calculée à partir des images exemples et contre-exemples. Notre architecture permet aussi de choisir dynamiquement le descripteur le plus adapté pour une notion ou un mot-clé consideré. En outre, le procédé de rétroaction (feedback) améliore les résultats obtenus tandis que l'annotation différée hors ligne (offline) élimine les restrictions de complexité et de temps de calcul, en fournissant une performance optimale lors de la phase de recherche qui exploite uniquement les mots-clé. De plus, des hiérarchies et combinaisons de mots-clés permettent l'écriture de requêtes de haut niveau sémantique. Le prototype expérimental ainsi développé doit faire l'objet d'une valorisation dans le cadre d'un contrat avec Oséo.

8. Conclusion

Afin de résoudre le problème relatif à l'extension des opérateurs morphologiques aux images couleur, principal objectif de cette thèse, les contributions suivantes ont été faites : une étude comparative des approches existantes, au travers d'applications en réduction de bruit et en classification de textures, qui a montré qu'aucune approche d'ordonnancement n'est appropriée pour tout type d'application, et donc qui a souligné l'importance du critère de flexibilité. Nous avons ainsi décidé de nous focaliser sur l'ordre lexicographique couplé avec l'espace de couleurs LST.

Ensuite, nous nous sommes concentrés sur les difficultés liées à cette configuration, et spécifiquement au traitement de la teinte, pour lequel une nouvelle approche basée sur l'utilisation de teintes multiples de référence a été introduite. Nous avons également étudié l'inconvénient majeur de l'ordre lexicographique, c'est-à-dire la forte priorisation de la première bande d'images, et avons proposé trois solutions. Finalement une définition de la transformée en tout-ou-rien vectorielle a été fournie. Néanmoins, le problème de la morphologie mathématique multivariée étant implicitement lié à celui de l'ordonnancement des vecteurs pour lequel aucune solution absolue n'existe, son application aux images couleur évoluera certainement au fur et a mesure de l'arrivée de nouveaux besoins applicatifs. Dans ce contexte, nos efforts principaux ont donc porté sur le concept de flexibilité.

Dans une seconde partie, plutôt relative aux les aspects applicatifs, nous avons dressé un panorama du domaine de la recherche d'images par le contenu, et après avoir souligné le défi majeur de la description du contenu sémantique, nous avons étudié le potentiel offert par le cadre morphologique pour la description du contenu des images couleur, et avons introduit plusieurs descripteurs globaux de couleur et de texture.

De plus, un système d'annotation et de recherche à base de mots-clés a été proposé, qui exploite les connaissances sémantiques d'un expert au moyen d'images exemples et contre exemples, utilisées pour déterminer de façon dynamique le descripteur optimal pour chaque mot-clé. Le domaine de la recherche d'images par le contenu étant particulièrement actif, le potentiel des opérateurs morphologiques dans ce contexte, souligné dans ce travail, offre un nombre important de perspectives telles que les points d'intérêt morphologique, la caractérisation de forme, etc.

Contents

Chapter 1

Introduction

This thesis deals with the extension of morphological operators to colour images. The approaches that are developed are then used in the context of content-based image retrieval.

1.1 Colour morphology

In terms of computer memory occupation, a colour image is equivalent to three times the size of the corresponding grey-level image with the same resolution and colour depth. However, many may argue that the amount of subjective information contained within a colour image may well be more than triple. In other words, a colour image can be sometimes worth more than three thousand words. Yet with the exception of cases where the independent processing of each colour channel is acceptable, the exploitation of this additional wealth of information has seldom been straightforward. As a matter of fact no matter what the digital image processing method under consideration is, the extension of various grey-level transformations and operators to colour images has often led to ambiguities; and mathematical morphology, a non-linear image processing framework, is no exception to this rule. The resolution of this problem is the main objective of this thesis, in other words to define an effective framework for the application of morphological operators to colour images.

One of the reasons leading to the complication of this matter is the very nature of colour. Even though the physical properties of light, resulting in different colour sensations, and the laws governing these relations have been scrutinised for many centuries, our understanding of the inner workings of the human colour vision system and of how the light entering through our eyes ends up providing our brain with colour images, is still incomplete. Therefore the initial difficulty of colour image analysis is the establishment of an effective colour representation.

Moreover, besides this inherent difficulty, the definition of morphological operators for colour images is further hindered by a fundamental theoretical requirement. Specifically, the morphological framework is based on complete lattices, thus in order to accommodate multivalued data such as colour images, a way of calculating the extrema of vector data is essential. Yet unlike scalars, there is no unambiguous way of ordering vectors. As a result, numerous vector ordering schemes coupled with a colour representation choice, have been employed for defining multivariate morphological operators. However, none of them has yet been widely accepted.

Among the proposed approaches, the thesis work of Hanbury (2002) [95], followed by that of Angulo (2003) [6] constitute particularly important contributions to this field. More precisely, Hanbury studied various colour space options for morphological operators, and opted for polar spaces, while concentrating especially on the morphological processing of hue. Then Angulo expanded his work, by experimenting with different ordering possibilities in the same type of colour space, where he introduced the *α-modulus lexicographical ordering* option. As a matter of fact, this part of the thesis constitutes in a way a continuation of their work. To explain, we first assert their choices of colour representation and colour ordering, and further develop some variations with the purpose of improving the performance of the resulting morphological operators.

More precisely, we have begun our study of this problem with the choice of a "suitable" colour space, which led to the question of defining the notion of suitability in this context. Next, a comparative study of various orderings along with multiple applications, showed that no single vector ordering approach can satisfy all practical needs equally well. Hence flexibility is a must, and consequently a suitable colour space is one that can accommodate it. That is why it was decided to endorse the luminance-saturation-hue based colour spaces, that are considered to represent the principles of the human colour vision system, and are thus appropriate for practical, intuitive and application dependent colour ordering configurations. Specifically, the LSH colour space has been chosen, that remedies certain serious saturation related faults of classic options such as HSV and HLS [95].

Having chosen a colour space, we then focused on vector ordering. In particular, we have concentrated on the lexicographical approach, due to its positive theoretical properties along with its flexibility to adapt to different practical needs. We then proposed multiple solutions to its main drawback: the extreme prioritisation of the first image channel, which in most cases leads to an ineffective exploitation of the available data. In detail, first the α-modulus lexicographical ordering was modified in order to support arbitrary priority distributions, hence making possible a finer transitional control during the lexicographical comparison cascade. In the particular case of colour images, this allows to better model the dependencies among luminance, saturation and hue. Moreover, an approach where the shift of transition is based on a marker image has also been studied. To explain, in this image specific approach, an auxiliary (i. e. marker) image is used in order to define the areas of the input where a particular channel is assumed to leave its lexicographical turn to the next. Furthermore, a third variation that has been named *α-trimmed lexicographical* has been introduced. This approach does not constitute an ordering and instead focuses on computing directly the extrema needed in order to define dilation and erosion, therefore leading to pseudo-morphological operators. Its user defined α argument is used in a way similar to the known *α-trimmed* filters, reducing the amount of vectors channel by channel, thus leading to unique extrema. The practical interest of all elaborated approaches has been illustrated through applications such as noise reduction and texture classification.

We additionally study the morphological processing of the hue component, which due to its periodic nature is susceptible to problems with scalar ordering. We show that the widely used approach based on a single reference hue value does not satisfy all practical needs, such as when more than one dominant hue values are present within the image. Therefore an alternative approach making use of multiple reference hues is presented. This part is concluded with the extension of the hit-or-miss transform to multivariate images as an application of its results. We provide a definition for a vector hit-or-miss transform (HMT) and study the theoretical requirements relative to the use of multivariate structuring functions. Furthermore, approaches such as rank filters and synthetic structuring elements, known for increasing the robustness of the hit-or-miss transform, are also elaborated in the multivariate case.

Following the development of the theoretical foundations for morphological colour image analysis, we then focus on its application to general purpose content-based colour image description and retrieval.

1.2 Content-based image retrieval

Multimedia content is an indispensable part of contemporary societies. The amount of visual data in particular produced by institutions as well as individuals is increasing constantly, thus leading to the creation of rich visual repositories, ranging in variety from home photo collections to astronomical image databases. Their effective exploitation however cannot be carried out with the same approches used for text-based databases, for which elaborate database management systems have been developed, since those methods have been designed for alphanumeric data. That is why new techniques specific to visual content become necessary, in order to organise index and retrieve information effectively.

This is the field of content-based image retrieval (CBIR), which matured rapidly by capturing the attention of both industry and research communities through its application spectrum. In detail, after the initial feasibility studies during the early 90's, the focus has gradually shifted to satisfying user needs, and in particular to the bridging of the *semantic gap*. Specifically, this last term refers to the gap between high-level image semantics (e. g. sunset, storm, etc) and low-level image features (e. g. colour, texture, shape, etc). It constitutes the main challenge of this field, which continues to be one of the most active of our time.

However, despite the long and successful use of mathematical morphology in several image processing problems, its application to this field dates back only a few years, a situation that could be partially justified by the lack of a widely accepted morphological framework for colour images. That is why besides its application to colour content description has been so far limited with highly specific cases, such as hematological cell classification [6], and colour textural analysis of wood [95], whereas its use towards general purpose CBIR had been left largely unexplored until very recently [267]. Above all, the main motivation for using morphological approaches in CBIR is to benefit from the framework's capacity to exploit spatial pixel relationships, a complementary property with respect to statistical approaches, and as confirmed by the results obtained so far, morphological operators can indeed contribute to this field.

In this second part of the thesis, we approached this problem first on the level of colour content description, where the processing of spatial along with spectral information, has long been a problem. More precisely, first an effective colour quantisation scheme has been chosen through experimentation, that has been followed by the introduction of colour specific granulometries, aiming to describe the size distribution of each quantised colour independently. Then, we focused on the concept of multi-resolution histograms, where we elaborate on two ways of integrating them with morphological operators. Specifically, we studied the effectiveness of multi-resolution histograms based on scale spaces obtained through colour levelings, a powerful connected morphological operator leading to flat zones, as well as through the watershed transform. Next, we proceeded to texture description, where the two main morphological texture characterisation tools, granulometry and covariance, have been combined through multiple structuring element variables, with the purpose of exploiting the complementary information that they provide.

Furthermore, having obtained some means of content description, we have continued to the implementation of a content-based description, annotation and retrieval architecture based on textual keywords. The use of keywords makes it possible for an expert to add her semantic knowledge into the system by means of adequately chosen example and counter example images, that are examined in order to choose the optimal descriptor for the keyword under consideration. The results are further refined through negative and positive feedback, whereas offline annotation eliminates most restrictions on descriptor complexity, while providing constant time retrieval performance. Additionally, keyword hierarchies as well as keyword combinations render the user capable of semantically higher level queries.

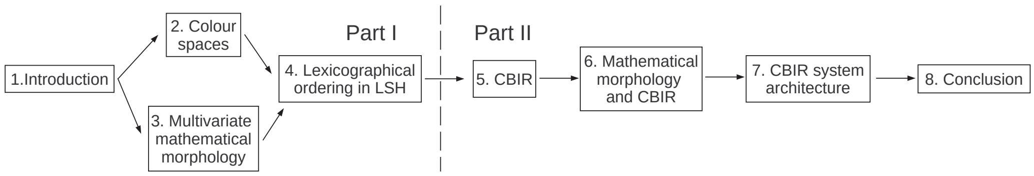


Figure 1.1: Chapter organisation.

1.3 Thesis organisation

The rest of this thesis is organised in two parts, representing respectively the theoretical and application oriented directions of our research objective. The first part starts with a presentation of human colour perception principles and various colour spaces along with their properties (Chapter 2). Chapter 3 continues to provide the state-of-the-art in multivariate morphology with focus on colour images, as well as a comparative study of the various approaches using multiple applications. Then in Chapter 4, we elaborate on our contribution to colour morphology, in particular we propose various extensions to lexicographical ordering, in combination with a phenomenal colour space, as well as a hue processing method and a vector HMT, thus concluding the first part. The second part begins with an overview of the field of CBIR, where the principal aspects of visual content retrieval are detailed (Chapter 5). Next, in Chapter 6, we present multiple morphological approaches for colour and texture description, that are tested against different known descriptors using various image collections. Moreover, the proposed descriptors are then put to practice in Chapter 7, where the developed retrieval system's design goals and architecture are detailed and retrieval examples are provided. Finally, Chapter 8 summarises our findings and their perspectives.

Part I

Morphological colour image analysis

Chapter 2

Colour spaces

2.1 Introduction

It has been argued that the human vision system has evolved to accommodate the concept of colour as a means to assess the ripeness of fruits [209]. Whatever the initial reason, colour is undoubtedly an intrinsic part of vision, that enriches our view of our environment in an unique way. Nevertheless, it remains an "illusion", in the sense that a colour represents nothing more than a subjective sensation that is associated with a certain combination of neural signals transmitted to our brain; for instance that which a healthy human being sees as red can be some other animal's purple.

Besides, our understanding of colour has evolved substantially in time, as a physical and as a psycho-physiological phenomenon. From a physical point of view, colour refers to our perception of visible light. Although studied since antiquity, breakthroughs in this field were realised first with the ground-breaking work of I. Newton in the late 17th century, when he experimentally showed that colour is not an inherent property of objects, but rather a result of their interaction with light. Two centuries later, through the work of T. Young, J. C. Maxwell and H. Helmholtz, the "trichromatic colour theory" was developed, stating that any colour can be simulated through the combination of different quantities of red, green and blue. On the other hand, during the same period, E. Hering established his "opponent colour theory", expressing colour in terms of three opposing couples, white-black, red-green and yellow-blue. It was finally in the 20th century, following additional psycho-physiological findings, that both approaches were validated by the "stage theory" of colour vision. In contrast to the physical aspects, physiological studies indicated since the 19th century that the advanced stages of human vision manipulate colour in terms of hue, saturation and brightness components [75].

Considering the aforementioned variety of dimensions to the notion of colour, a rich number of colour specification and visualisation methods, or *colour spaces* are nowadays available. While some were designed and developed with a certain family of applications in mind, others were conceived as general purpose colour representations. Each color space has its own set of desirable properties as well as drawbacks for purposes of colour image processing and computer vision. Prior to making our colour space choice, we find it pertinent to present an overview of the principal colour spaces, with focus on those used in, but not limited to, computer graphics. For further information on colour theory and colour spaces the reader may consult Refs. [71, 203, 214, 266]. In detail, this chapter continues with a brief summary of the physiological principles of human colour perception, followed by the presentation of different colour representations, categorised as RGB, perceptually uniform, application specific and phenomenal.

2.2 Principles of human colour perception

The perception of colour in humans and more generally in primates, is made possible through the presence of photoreceptor cells, called *cones*, on the retinal wall of the eye. Cones are sensitive to different wavelength groups of visible light, which represents roughly the interval between $380nm$ and $740nm$ of the electromagnetic spectrum. Thus, the combination of their neural responses to the reception of different wavelengths of light is what initiates the extremely complicated process of colour vision.

From a physiological point of view, cones are distinguished into three types, L, M and S, according to their spectral sensitivity. To explain, L-cones are sensitive to longer wavelengths of the visible light spectrum, representing the red colour, whereas M-cones respond to medium wavelengths, corresponding to green. S-cones on the other hand, are sensitive only to shorter wavelengths that correspond to blue and violet light. It is due to the presence of only these three types of cones, that humans are considered trichromatic, hence being able to perceive through their combinations, roughly 10 million different colours [127]. However, cones represent a small minority among the photoreceptor cells within the eye, numbering around 6-7 million and concentrated mainly at the *fovea*, a central region of the retina of extremely high visual acuity. A second type of photoreceptor cell, called *rods*, occupy the majority of the retinal surface, numbering around 120 million, with a more homogeneous distribution, everywhere outside the fovea. Different than cones, rods are active only in conditions of dim light, making them responsible for vision in dark environments (i. e. scotopic vision), whereas cones become responsive only when the amount of light entering the eye is relatively strong (i. e. photopic vision).

Based on this physiological information, the contemporary “stage theory” of colour vision, constitutes a combination of two previous theories: the “trichromatic colour” and “opponent colour” theories. Specifically, the first was developed in the 19th century, initially by T. Young and H. Helmholtz, who through experimental observations argued that the human eye possesses only three types of receptors sensitive to three primary colours, and by combinations of which all colours are perceived. Later J. C. Maxwell further expanded their work and determined red, yellow-green and blue as primary colours, that he used in order to realise the first colour photograph. However, this theory was still unable to explain various visual phenomena. Therefore, at approximately the same period E. Hering presented his opponent colour theory supported by experimental results, according to which yellow is not perceived as a combination of red and green, but rather constitutes an elementary colour opposing blue, just like red opposes green. This “antagonistic” view of colours helped explain why we do not see any “reddish-green” or “yellowish-blue”. These theories were considered as contradicting until the mid-20th century, when they were reconciled through the work of L. Hurvich and D. Jameson.

The contemporary colour vision theory (called *opponent theory of colour vision* or *stage theory*), is illustrated in Fig. 2.1, and encapsulates both previous theories. As a matter of fact, the trichromatic point of view, explains the initial reception stage of light by the retinal wall, where indeed three types of colour sensitive photoreceptor cells are located. At this point, in conformance with the opponent colour theory the signals produced by these cells

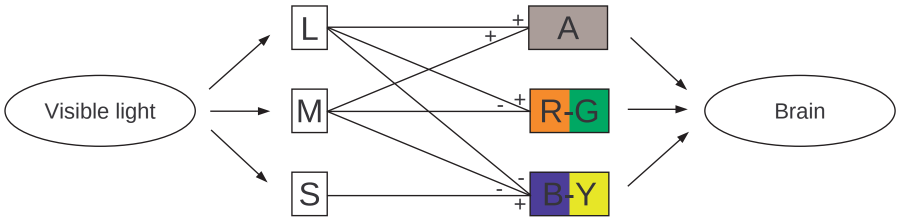


Figure 2.1: Illustration of the stage theory of colour vision.

pass from a stage of neural processing, where they are re-encoded as opposing couples prior to transmission to the brain's visual cortex. More precisely, one achromatic (white-black, $L + M$) and two chromatic (red-green, $L - M$ and blue-yellow, $S - M - L$) outputs are generated [80]. This decorrelating transformation is believed to allow a more efficient and robust signal transmission. It is also known that the achromatic channel carries a relatively higher quantity of spatial detail with respect to the other two channels, a property which is being effectively exploited by edge detectors and by the broadcasting systems that allocate a larger bandwidth to this channel. As a matter of fact, the achromatic channel is often considered sufficient for the recognition of most objects, while chrominance information, which is absent at low illumination levels represents a rather auxiliary component, thus forming two separate qualities of human vision.

Although the initial stages of colour vision have been fairly thoroughly studied, much less is known about the later stages. Nevertheless, psycho-physiological experiments realised as early as the 19th century have given indications on the intuitiveness, as far as human observers are concerned, of the notions of hue, saturation and brightness, as means to describe colour. Furthermore, recently proposed models show the feasibility of the derivation of these components through non-linear neural mechanisms, therefore increasing their physiological plausibility [25]. Having provided an insight into the inner workings of human colour perception, we now proceed to an overview of the different methods that have been developed for the specification and visualisation of colour.

2.3 A colour standard

In the history of colour study, the first standardisation attempt of colour notation and specification was realised by the *International Commission on Illumination* (CIE, *Commission International de l'Eclairage*) in 1931 [47]. This commission, which is nowadays considered the international authority on illumination and colour related topics, defined a set of standards concerning the three elements leading to the perception of colour: the source of illumination, the illuminated object and the observer. One of CIE's initial accomplishments was the definition of the first mathematically designed colour space: the *CIE 1931 XYZ colour space*, which is based on human visual perception and continues to serve as a basis for colorimetric processes even today. Its design purpose was to produce a colour space, making possible the precise exchange and communication of colour information, once the aforementioned parameters had been well defined.

Furthermore they provided standard definitions for the various aspects of colour. Colour intensity for instance was denoted by three concepts: *brightness*, *luminance* and *lightness* [48].

While brightness is a subjective notion, measuring how much light an area emits, luminance is a physical quantity, measuring the luminous intensity per surface unit, and is defined as a weighted sum of the three primary colours according to the spectral sensitivities of the human eye. Lightness on the other hand, takes into account additionally the non-linear response of the human eye to luminance.

This last point merits a more detailed explanation as it is has a certain effect on colour operations. Specifically, the relation between the voltages that encode the pixel intensities and the photons produced by the electron gun of cathode ray tube (CRT) display systems, is non-linear and more precisely a power function (the almost inverse relation of which coincidentally represents the lightness sensitivity of human vision). That is why, in order to compensate for it, recorded intensity values are "corrected" using an approximation of the inverse of the aforementioned power function (i. e. gamma correction), usually applied at the acquisition device. According to Ref. [244], the colour values resulting from this operation, that are besides manipulated by most imaging algorithims, are incompatible with arithmetic operations. Nevertheless, the correction function is regular enough to be well approximated by limited expansions, hence making it possible all the same to develop vector formalisms with the corrected values, even though they are physically sound only for their uncorrected versions [244].

Moreover, CIE defined a set of reference *colour-matching* functions for the perception of red, green and blue by a standard observer, that have led subsequently to the definition of the RGB colour space.

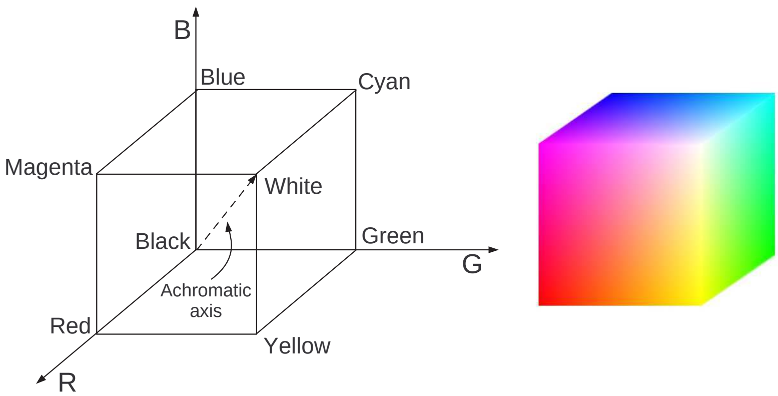


Figure 2.2: The RGB colour cube.

2.4 RGB

Red, Green, Blue (RGB) is by far the most widely used colour space in imaging environments. It can be found in televisions, computer monitors and digital cameras among others. Theoretically, it is based on the concept of trichromaticity and on the reference colour-matching functions defined by CIE, therefore it represents colours as combinations of the three primaries. Specifically, it is considered as an *additive* colour model. In additive colour reproduction, the mixing of primary colours with the purpose of obtaining secondary ones, is realised on the basis of light emission, conversely to subtractive models where light reflection is taken as reference. Consequently, combining for instance red and green leads to yellow (R = max, G = max, B = 0), whereas red and blue gives magenta (R = max, G = 0, B = max). Black is considered as the absence of light, hence no primary colour is necessary for its production (R = G = B = 0). White on the other hand is equivalent to the presence of all primary colours in maximal amounts (R = G = B = max).

The classical way of visualising the RGB space is by means of a cube, as shown in Fig. 2.2, where each primary colour is denoted by an axis, and every point within the cube represents a colour in the form of a RGB triplet. Moreover, the diagonal connecting the black and white corners corresponds to the achromatic axis containing the shades of grey.

Despite its widespread use however, RGB suffers from multiple inconveniences. First, with the exception of the eight cube corners, it is intuitively difficult to specify arbitrary colours by means of the RGB cube. Furthermore, the distances between points residing inside the cube are not proportional to their perceptual differences; in other words, RGB is perceptually non-uniform. Additionally, the exact values of the R, G and B components of a given colour depend on the spectral sensitivity functions of the acquisition device under consideration, hence rendering this space device-dependent. And finally, the three components of RGB are highly correlated, specifically, $\rho_{BR} = 0.78$, $\rho_{RG} = 0.98$ and $\rho_{GB} = 0.94$ [203]. Due to these drawbacks, RGB has been considered in general as an unsuitable colour space for digital image processing.

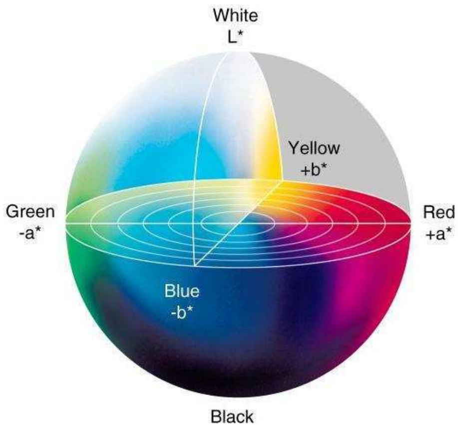


Figure 2.3: Spatial schema of CIELAB (http://www.colorspan.com).

2.5 Perceptually uniform colour spaces

Perceptual uniformity, the property which for a change of colour value induces a perceptually proportional change, has been a long desired property of colour spaces. As this invaluable property for computing colour differences was not provided by the initial XYZ colour space, in 1976 CIE adopted both *CIELAB* (or $L^*a^*b^*$) and *CIELUV* (or $L^*u^*v^*$), when no consensus could be achieved over one or the other. Both of these spaces are considered almost perceptually uniform or linear (linear for small values), and derive from XYZ, so given an image produced by an acquisition device in RGB, it is necessary to be able to convert it into and from XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = A \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = A^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.5.1}$$

where the RGB values are normalised to $[0, 1]$, and the matrix A used with this purpose depends on the reference white point corresponding to the illuminant under consideration. For instance, for the widely popular *D65* or *daylight illuminant*, it becomes:

$$A_{D65} = \begin{bmatrix} 0.412 & 0.358 & 0.180 \\ 0.212 & 0.716 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \quad \text{and} \quad A_{D65}^{-1} = \begin{bmatrix} 3.240 & -1.537 & -0.499 \\ -0.969 & 1.876 & 0.042 \\ 0.056 & -0.204 & 1.057 \end{bmatrix} \tag{2.5.2}$$

As both L*a*b* and L*u*v* follow the same principles, in this section we focus solely on L*a*b*. Basically, it represents a formulation based directly on the stage theory of colour vision, elaborated in Sec. 2.2, with the $L^* \in [0, 100]$ component denoting lightness, a^* the red-green and b^* the yellow-blue opposition. Thus, a^* and b^* possess negative values when the colour they represent is close to respectively green and blue, and positive values when it is closer to respectively red and yellow (Fig. 2.3). Given the XYZ triplet, the corresponding L*a*b* representation may be obtained as follows:

$$\begin{aligned} L^* &= 116 f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* &= 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \\ f(t) &= \begin{cases} t^{1/3} & \text{if } t > 0.008856 \\ 7.787t + \frac{16}{116} & \text{otherwise} \end{cases} \end{aligned} \tag{2.5.3}$$

where X_n, Y_n, Z_n denote the XYZ triplet representing the white point ($R = G = B = 1$), which can be obtained from Eq. (2.5.2). Consequently, L*a*b* inherits the white point reference used in the preceding XYZ expression. The inverse transformation that results in a XYZ colour representation is given by:

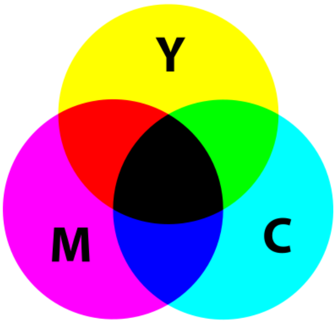
$$\begin{aligned} Y_n &= \begin{cases} Y_n \left[\frac{L^*+16}{116}\right]^3 & \text{if } L^* > 7.9996 \\ Y_n \left[\frac{L^*}{903.3}\right] & \text{otherwise} \end{cases} \\ X &= X_n \left[\left(\frac{Y}{Y_n}\right)^{1/3} + \frac{a^*}{500}\right]^3 \\ Z &= Z_n \left[\left(\frac{X}{X_n}\right)^{1/3} - \frac{b^*}{200}\right]^3 \end{aligned} \tag{2.5.4}$$

In this three-dimensional colour space colour distances become proportional to their perceptual difference. Given two colours $\mathbf{c}_1 = (L_1^*, a_1^*, b_1^*)$ and $\mathbf{c}_2 = (L_2^*, a_2^*, b_2^*)$, their sensation difference is computed as:

$$\Delta E_{Lab}^* = \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2} \tag{2.5.5}$$

where a difference of $\Delta E^*_{Lab} = 2.3$ is considered barely noticeable.

The capacity to compute perceptual colour differences is considered to be one of the major advantages of L*a*b*, that has been exploited in various colour image applications, requiring device-independence and/or perceptual uniformity. Examples include colour image classification [3], segmentation through clustering [163], colour object tracking [182], colour description for content-based image retrieval [265], as well as the extension of mathematical morphology to colour data [104]. On the other hand, it has been also recently argued that even though perceptual uniformity is a highly desirable property for colour definition and description, its use in the context of image processing is limited, since images are not processed in terms of "perceived values" [244]. Moreover, one cannot benefit from the property of device independence in cases where no a priori information is available on the acquisition conditions of the images under consideration, which is more than often the case in practice, as far as general purpose content-based image retrieval is concerned. Additionally, the transformations to and from RGB become computationally complex with the intermediate role of XYZ, hence the practical interest of L*a*b* has been considerably limited in real-time applications.

a) 

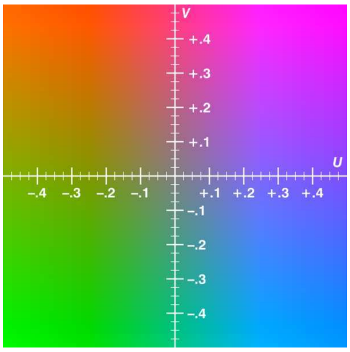
b) 

Figure 2.4: a) an example of subtractive colour mixing, b) Example of the U-V plane for Y = 0.5.

2.6 Application specific colour spaces

Certain colour spaces that have been developed since the pioneering work of CIE have been conceived with focused application areas, that have particular needs as far as colour specification and visualisation is concerned. These application areas include mainly printing and TV transmission.

In printing, since the printed medium cannot act as an illumination source, unlike for instance computer monitors, the additive mixture principle cannot be employed. Instead the subtractive colour reproduction model is used, which is based on the principle of reflected light. To explain, an object that is illuminated and perceived by an observer as green, means that its surface absorbs the other two primaries, red and blue, and reflects the green portion of visible light. Consequently, in order to reproduce the colour green in printing, it is sufficient not to use the colour absorbing green, which is green's complement: magenta; thus leading to a combination of cyan and yellow ink, the colours that absorb respectively the red and blue portions of light. Or, in order to obtain the colour black, in other words the absence of colour, the received light must be absorbed in its entirety by the illuminated object surface, meaning that all three primaries must be absorbed, so the complement of both R, G and B is used: cyan, magenta and yellow. This is the basic subtractive mixing principle behind the *CMYK* (*Cyan, Magenta, Yellow, Black*) colour space, used widely in printing applications (Fig. 2.4a) [203]. In addition, CMYK is device dependent, unintuitive and perceptually non-uniform, thus it is of little practical use for colour image processing, besides the printing field.

As far as analog TV transmissions and colour video standards are concerned, at the time of design of a colour space suitable for this type of use, the support for non-colour TV sets was crucial. Consequently, the separation of achromatic and chromatic information became essential, leading to the *Y'UV* space, developed for the PAL composite colour video standard. Y' stands for gamma-corrected luminance [214] and the other two components for colour related data (Fig. 2.4b). Thus, non-colour TV sets can use the same broadcast effectively by simply ignoring the chromatic information. In addition Y'UV is based on XYZ, and uses the D65 white point, while its conversion matrices are given by:

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \tag{2.6.1}$$

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.140 \\ 1 & -0.395 & -0.581 \\ 1 & 2.032 & 0 \end{bmatrix} \begin{bmatrix} Y' \\ U \\ V \end{bmatrix} \tag{2.6.2}$$

where R', G' and B' stand for gamma-corrected R, G and B. Due to its initial design purpose, the use of Y'UV in terms of image processing is concentrated on computer vision related applications [138], especially those making use of Y'UV based hardware. Variations of Y'UV include Y'IQ, YCbCr and YPbPr.

2.7 Phenomenal colour spaces

Phenomenal or *polar* colour spaces are those that express colours in intuitive terms as far as human observers are concerned. This type of colour representation is mainly due to the pioneering work of A. H. Munsell, who carried out rigorous experiments on the colour perception of humans [185]. He was the first to separate colour into the following notions:

- *hue*, which represents the dominant wavelength of a colour, described as "green","blue" etc. It is an angular, modulo 2π value $H \in [0, 2\pi] = [0°, 360°]$, with $0°$ corresponding to red.

- *saturation* ($S \in [0, 1]$), that describes the purity of a colour. A fully saturated colour appears vivid, whereas zero saturation transforms it into a shade of grey. Thus saturation can be also considered as the distance from the achromatic axis.

- *brightness*, *luminance* or *value* (L), that denotes the component describing the sensation associated with the amount of emitted light, or in other words the "amplitude" of the colour under consideration. The difference of brightness and luminance is detailed in Sec. 2.3. Likewise to saturation, it is often normalised to $[0, 1]$.

Polar spaces in their majority are obtained through direct transformations from RGB, hence being perceptually non-uniform and device dependent [213]. An additional disadvantage that further hinders their use with a purpose besides colour specification, is the angular nature of hue, which through modulo based operations leads to visual discontinuities. Moreover, the most widely used representatives of this category, HSV and HLS, as it was thoroughly detailed in Refs. [95, 99], suffer from major structural inconveniences. For these reasons, phenomenal spaces have long been viewed as unsuitable for colour image processing [213].

2.7.1 HSV and HLS

Both HSV and HLS are obtained through non-linear transformations from RGB, with their hue components being defined identically. In particular, given a RGB space normalised to $[0,1]$, the common approach for transforming it into HSV is as follows:

$$
\begin{aligned}
V_{HSV} &= \max\{R,G,B\} \\
S_{HSV} &= \begin{cases} \frac{\max\{R,G,B\}-\min\{R,G,B\}}{\max\{R,G,B\}} \text{ if } \max\{R,G,B\} > 0 \\ \text{undefined otherwise} \end{cases} \\
H_{HSV} &= \begin{cases} \text{undefined if } S_{HSV} = 0 \\ 60^\circ \times \frac{G-B}{\max\{R,G,B\}-\min\{R,G,B\}} + 0^\circ \text{ if } \max\{R,G,B\} = R \text{ and } G \geq B \\ 60^\circ \times \frac{G-B}{\max\{R,G,B\}-\min\{R,G,B\}} + 360^\circ \text{ if } \max\{R,G,B\} = R \text{ and } G < B \\ 60^\circ \times \frac{B-R}{\max\{R,G,B\}-\min\{R,G,B\}} + 120^\circ \text{ if } \max\{R,G,B\} = G \\ 60^\circ \times \frac{R-G}{\max\{R,G,B\}-\min\{R,G,B\}} + 240^\circ \text{ if } \max\{R,G,B\} = B \end{cases}
\end{aligned}
\tag{2.7.1}
$$

where the undefined saturation case corresponds to the intuitive situation of zero brightness, as is that of hue, which cannot be defined if it lacks all saturation. Conventionally, in these cases the value of 0 is used. As far as HLS is concerned, it can be obtained through the following transformations:

$$
\begin{aligned}
L_{HLS} &= \frac{1}{2}(\max\{R,G,B\} + \min\{R,G,B\}) \\
S_{HLS} &= \begin{cases} 0 \text{ if } \max\{R,G,B\} = \min\{R,G,B\} \\ \frac{\max\{R,G,B\}-\min\{R,G,B\}}{\max\{R,G,B\}+\min\{R,G,B\}} \text{ if } L_{HLS} \leq 0.5 \\ \frac{\max\{R,G,B\}-\min\{R,G,B\}}{2-(\max\{R,G,B\}+\min\{R,G,B\})} \text{ if } L_{HLS} \geq 0.5 \end{cases} \\
H_{HLS} &= H_{HSV}
\end{aligned}
\tag{2.7.2}
$$

from which various other representations may be obtained by modifying for instance the brightness expression (e. g. HSI). Furthermore, HSV and HLS are characterised by their cone and double-cone shapes respectively, as illustrated in Fig. 2.5. By varying the length of the saturation axis along that of luminance, they take into account the non-constant sensitivity of human vision to colour, depending on the amount of light available. As explained earlier in Sec. 2.2, an increase of the amount of light leads to the transition of photo-reception activity from rods to cones. Hence, HSV offers a maximum saturation range ($S \in [0,1]$) for maximum luminance ($L = 1$), while HLS does the same for $L = 0.5$. However, due to these shapes, it becomes possible to specify colour coordinates that lie outside the conical shapes, thus representing invalid colours, for example non zero saturation values for zero luminance. That is why during the 70's, validity controls were required in graphic applications during colour specifications. In order to avoid this complex (for then) task, these conical spaces

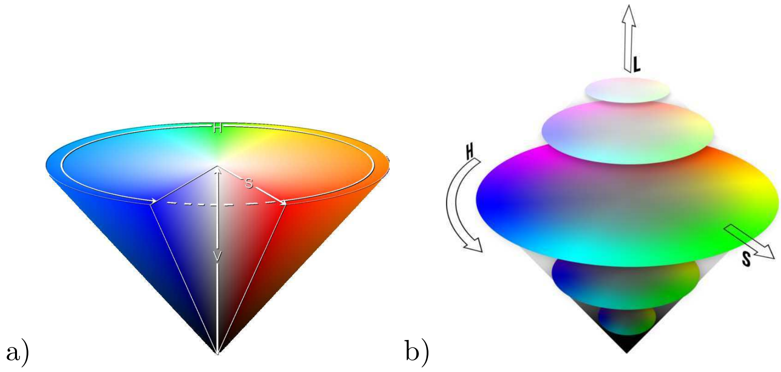


Figure 2.5: a) Conical HSV space, b) Biconical HLS space.

were modified into cylinders, through the division of the saturation component by luminance. This modification was observed by Hanbury and Serra [95, 99, 243], who remarked that the widely used transformations of Eqs. (2.7.1) and (2.7.2), lead in fact to cylinder-shaped colour spaces. They further carried out an extensive study of these spaces, and observed multiple inconsistencies, due principally to their cylindrical form.

First, by dividing saturation with luminance it becomes possible to have the counter intuitive situation where dark pixels possess extreme saturation values. Consider for instance the normalised RGB triplet $(0.01, 0.0, 0.0)$ representing a black-like colour, which through Eq. (2.7.1) results in a maximal saturation of $S_{HSV} = 1.0$. Furthermore, one cannot use a constant saturation threshold in cylinder shaped spaces as it is often practised, since the points corresponding to equal saturation do not form a straight line. Moreover, phenomenal spaces are inherently assumed to accomodate the separation of chromatic and achromatic information, however it can be easily remarked that saturation is in fact dependent on the chosen brightness function:

$$S_{HSV} = 1 - \frac{\min\{R,G,B\}}{V_{HSV}} \quad \text{and} \quad S_{HLS} = \begin{cases} \frac{\max\{R,G,B\}-\min\{R,G,B\}}{2L_{HLS}} & \text{if } L_{HLS} \leq 0.5 \\ \frac{\max\{R,G,B\}-\min\{R,G,B\}}{2(1-L_{HLS})} & \text{if } L_{HLS} \geq 0.5 \end{cases} \tag{2.7.3}$$

It is further shown in Ref. [95] that in both HSV and HLS spaces, neither brightness nor saturation are norms, hence hindering even basic colour operations in these spaces, such as colour addition and averaging.

2.7.2 An improved phenomenal space

Given the drawbacks of HSV and HLS, certain minimal prerequisites were established, in order to formulate a phenomenal space deriving from the RGB cube, suitable for colour image processing. These prerequisites are [95]:

- the independence of saturation from brightness,
- equipping the notions of brightness and saturation with norms,
- and the inversibility of this space back to RGB.

Norm	Brightness (L)	Saturation (S)
$L2$	$\sqrt{\frac{R^2+G^2+B^2}{3}}$	$\frac{3}{2}\sqrt{(2R-G-B)^2+(2G-B-R)^2+(2B-G-R)^2}$
$L1$	$\frac{R+G+B}{3}$	$\frac{3}{4}(\lvert R-L\rvert+\lvert G-L\rvert+\lvert B-L\rvert)$
$max-min$	Y	$\max\{R,G,B\}-\min\{R,G,B\}$

Table 2.1: The brightness and saturation expressions that have been developed for three different norms [95]. $Y = 0.213R + 0.715G + 0.072B$ represents the recommended luminance formulation of high definition television.

With these conditions to verify, the appropriate transformations were formulated, where the initial prerequisite is satisfied by removing the division of saturation by the brightness component:

$$S'_{HSV} = S_{HSV}V_{HSV} \text{ and } S'_{HLS} = S_{HLS}\left(1-2\left|\frac{1}{2}-L_{HLS}\right|\right) \tag{2.7.4}$$

hence effectively achieving the conical and biconical shapes respectively for HSV and HLS. While for the second prerequisite, multiple norm options have been elaborated: L_2, L_1 and the semi-norm $max - min$. The brightness and saturation expressions corresponding to each norm are given in Table 2.1. As all three of them are theoretically valid, this choice became purely a question of practical interest. Besides, this problem was undertaken later by Angulo and Serra [6, 12], who realised a comparative study in the context of colour image segmentation. According to their observations, a luminance expression representing the perceptual characteristics of the human eye is not of particular use for quantitative colour image processing, since the focus is on the *object under study* and not on the *perceived image*. That is why they have opted for the most symmetrical luminance expression, the average of RGB components which corresponds to the L_1 norm. Besides, this norm also suits the linear model representing the additivity of the energy quantities leading to the perception of colour, conversely to the L_2 norm.

The biconical phenomenal space based on the L_1 norm for both brightness and saturation, has been characterised as LSH [8] and its transformation from RGB using a simplified hue expression is as follows [243]:

$$\begin{aligned} L_{LSH} &= \frac{1}{3}(\max + \text{med} + \min) \\ S_{LSH} &= \begin{cases} \frac{3}{2}(\max - L_{LSH}) & \text{if } L_{LSH} \geq \text{med} \\ \frac{3}{2}(L_{LSH} - \min) & \text{if } L_{LSH} \leq \text{med} \end{cases} \\ H_{LSH} &= k\left[\lambda + \frac{1}{2} - (-1)^{\lambda}\left(\frac{\max + \min - 2\,\text{med}}{2S_{LSH}}\right)\right] \end{aligned} \tag{2.7.5}$$

where max, med and min denote respectively the maximum, median and minimum value of the transformed RGB triplet. k is the angle unit ($\pi/3$ for radians and 42 for 256 levels), whereas λ is set as:

$$\begin{array}{lll} 0 \text{ if } r > g \geq b & 1 \text{ if } g \geq r > b & 2 \text{ if } g > b \geq r \\ 3 \text{ if } b \geq g > r & 4 \text{ if } b > r \geq g & 5 \text{ if } r \geq b > g \end{array}$$

A detailed explanation of the inverse transformation can be found in Refs. [6, 105]. In short, it consists in using first the relation $H = (\lambda + \phi)k$ with $\phi \in [0, 1)$, for obtaining the order of R, G and B through the value of λ, as well as the value of ϕ. Then, depending on whether $(-1)^\lambda(\phi - 1/2) \leq 0$ or $(-1)^\lambda(\phi - 1/2) > 0$, the respective of the following two systems is used:

$$\begin{aligned} \max &= L_{LSH} + \frac{2}{3}S_{LSH} \\ \text{med} &= L_{LSH} - \frac{1}{3}(-1)^\lambda S_{LSH} + \frac{2}{3}(-1)^\lambda S_{LSH}\phi \\ \min &= L_{LSH} - \frac{1}{3}S_{LSH} - \frac{2}{3}S_{LSH}\,\text{rem}(\lambda, 2) - \frac{2}{3}(-1)^\lambda S_{LSH}\phi \end{aligned} \tag{2.7.6}$$

or

$$\begin{aligned} \max &= L_{LSH} + \frac{1}{3}S_{LSH} + \frac{2}{3}S_{LSH}\,\text{rem}(\lambda + 1, 2) - \frac{2}{3}(-1)^\lambda S_{LSH}\phi \\ \text{med} &= L_{LSH} - \frac{1}{3}(-1)^\lambda S_{LSH} + \frac{2}{3}(-1)^\lambda S_{LSH}\phi \\ \min &= L_{LSH} - \frac{2}{3}S_{LSH} \end{aligned} \tag{2.7.7}$$

where $\text{rem}(a, b)$ is the remainder of the division of a by b, and the resulting R, G and B values are set according to the order corresponding to the value of λ. Hence LSH, along with the conic and biconical representations of HSV and HLS, represent important improvements over the original definitions, that render them suitable for colour image processing applications where device independence and perceptual uniformity are not essential. If however these two elements are necessary, then the L*C*H* colour space is a more adequate choice.

2.7.3 L*C*H*

The L*C*H* colour space was developed in order to address the issue of device dependence and perceptual uniformity in the case of phenomenal spaces. Simply put, it is the polar representation of the L*a*b* space, where colours are represented in terms of lightness, *chroma* (i. e. a saturation related notion), and hue. Specifically, the last two components are given by:

$$\begin{aligned} C^* &= \sqrt{(\text{a}^*)^2 + (\text{b}^*)^2} \\ H^* &= \arctan\left(\frac{\text{b}^*}{\text{a}^*}\right) \end{aligned} \tag{2.7.8}$$

As to the perceptual colour difference in the polar representation, given $\mathbf{c}_1 = (L_1^*, C_1^*, H_1^*)$ and $\mathbf{c}_2 = (L_2^*, C_2^*, H_2^*)$, it is computed as follows:

$$\Delta E_{LCH}^* = \sqrt{(L_1^* - L_2^*)^2 + C_1^* + C_2^* - 2C_1^*C_2^*\cos(H_1^* - H_2^*)} \tag{2.7.9}$$

Consequently, L*C*H* represents one of the most interesting options of phenomenal spaces, that combines perceptual uniformity with the colour representation intuitiveness of this category.

	Intuitiveness	Transformation speed	Perceptual uniformity	Device independence
RGB	-	+	-	-
$L^*a^*b^*$	-	-	+	+
$L^*C^*H^*$	+	-	+	+
LSH	+	+	-	-

Table 2.2: An overview of the properties of some of the colour spaces presented in this chapter.

2.8 Conclusion

This chapter has provided a brief overview of the principles of human colour perception and of the principal colour spaces used in various fields, with focus on computer graphics. Their properties, as far as colour image processing is concerned have been elaborated and the main points are summarised in Table 2.2. More precisely, considering the numerous disadvantages of RGB, it is obviously not a viable option. However, as the property distribution diversifies, no optimal option is evident. A rough categorisation can be made based on perceptual uniformity, the desirability of which in image processing is questioned, and device independence, which are only satisfied by CIE based colour spaces, and represent the only available options for applications requiring these characteristics. Nevertheless, their transformation complexity obviates to a significant degree their practical use. Phenomenal spaces are intuitive, offer good options, though they are all handicapped with uncertainty in the processing of the angular hue value.

In conclusion, the "optimal" colour space choice clearly depends on the needs of the application under consideration. In which case, the next logical question to ask is: what are the needs of mathematical morphology, as far as the colour space in use is concerned? This question is answered in the next chapter.

Chapter 3

Multivariate mathematical morphology

3.1 Introduction

Having provided an insight into colour spaces, in this chapter we focus on the overlying application: colour morphology, which constitutes a subset of multivariate mathematical morphology, and establish the notions required for its definition.

More precisely, the mathematical morphology (MM) theory, founded by G. Matheron [171] and J. Serra [240, 241], is a powerful image analysis framework, nowadays fully developed for both binary and grey-level images. Its popularity in the image processing community is mainly due to its rigorous mathematical foundation as well as its inherent ability to exploit the spatial relationships of pixels. The morphological framework provides a rich set of tools able to perform from the simplest to the most demanding tasks: noise reduction, edge detection, segmentation, texture and shape analysis, etc. As a methodology, it has been applied to almost all application areas dealing with digital image processing [250]. Consequently, it was only a matter of time before attempting to extend the same concepts to colour and more generally to multivalued images, i. e. images with more than one channels.

Unfortunately, this extension is not straightforward. Specifically, the morphological framework is based on complete lattices [110], thus in order to accommodate multivalued images, a way of calculating the extrema of vector data is essential. Yet unlike scalars, there is no unambiguous way of ordering vectors. And besides, most of the known vector ordering schemes have already been employed for defining multivariate morphological operators. However, none of them has yet been widely accepted.

This chapter initially provides an overview of the theoretical background of multivariate morphology (Sec. 3.2), and then continues to elaborate on the existing approaches used for extending the fundamental operators of MM to multivariate data (Sec. 3.3); and that is why cases limited only to specific operators have been omitted (e. g. colour image segmentation). As our primary purpose is to present the state-of-the-art in colour morphology, we focus particularly on cases developed with colour images in mind. Nevertheless, the scope of the survey is not limited to colour specific vector ordering approaches, since theoretically an ordering is independent from the represented context (e. g. a colour, a multispectral pixel value, etc). That is why morphological solutions that have been developed for non-colour data, but can however be trivially applied to this case, are also included. Due to their rich diversity and overwhelming number they are examined according to their methodology of ordering their multivalued input.

Moreover, we then further proceed to an application oriented comparative study of the major approaches (Sec. 3.4). We conduct tests in noise reduction and texture classification, with the goal of both comparing the performance of different vector ordering schemes as basis for morphological operators, and for asserting their theoretical properties with the help of experimental results. Based on the obtained data, as well as on the colour space related information presented in the previous chapter, we conclude by deciding which vector ordering and colour space to use, in order to define colour morphological operators in the sequel of this thesis (Sec. 3.5).

3.2 Extension of MM to multivariate data

In this section, we recall briefly the theoretical concepts behind the extension of morphological operators to multivariate images. For an in-depth study of the theory behind multivariate mathematical morphology the reader can refer to Refs. [86, 242].

3.2.1 Orderings

As the concept of order plays a central role in this field, we start by recalling the relative definitions. A binary relation $\mathcal{R}$ on a set $\mathcal{S}$ is called:

- *reflexive* if $x \,\mathcal{R}\, x, \;\; \forall\, x \in \mathcal{S}$
- *anti-symmetric* if $x \,\mathcal{R}\, y$ and $y \,\mathcal{R}\, x \Rightarrow x = y, \;\; \forall\, x, y \in \mathcal{S}$
- *transitive* if $x \,\mathcal{R}\, y$ and $y \,\mathcal{R}\, w \Rightarrow x \,\mathcal{R}\, w, \;\; \forall\, x, y, w \in \mathcal{S}$
- *total* if $x \,\mathcal{R}\, y$ or $y \,\mathcal{R}\, x, \;\; \forall\, x, y \in \mathcal{S}$

A binary relation $\leq$ that is reflexive and transitive is called a *pre-ordering* (or *quasi-ordering*); if the anti-symmetry constraint is also met, it becomes an *ordering*. If additionally the totality statement holds for $\leq$, it is denoted as *total*, if not *partial*.

3.2.2 Complete lattices and multivariate morphology

The complete lattice theory is widely accepted as the appropriate algebraic basis for mathematical morphology. Besides unifying the approaches previously employed in binary and grey-level morphology, it also makes it possible to generalise the fundamental concepts of morphological operators to a wider variety of image types and situations. Extensive details on the lattice based description of the MM theory can be found in Refs. [110, 112, 225, 229, 241, 242]. As a remark, we should also add that the scope of MM has also been extended to complete semi-lattices that are more general than complete lattices [111, 131, 132].

Specifically, a *complete lattice* $\mathcal{L}$ is a non empty set equipped with a partial ordering $\leq$, such that every non-empty subset $\mathcal{P}$ of $\mathcal{L}$ has a greatest lower bound $\bigwedge \mathcal{P}$, called *infimum*, and a least upper bound $\bigvee \mathcal{P}$, called *supremum*. In this context, images are modelled by functions mapping their domain space $\mathcal{E}$, an arbitrary non empty set that is an abelian group with respect to $+$, into a complete lattice $\mathcal{T}$ (with $\top$ and $\perp$ respectively its greatest and least elements), defining the set of possible "grey values". Moreover, if $\mathcal{F}$ represents the set of functions $f : \mathcal{E} \to \mathcal{T}$, then for the partial ordering:

$$f, g : \mathcal{E} \to \mathcal{T}, \;\; f \leq g \Leftrightarrow \forall\, x \in \mathcal{E}, \;\; f(x) \leq g(x) \tag{3.2.1}$$

$\mathcal{F}$ also forms a complete lattice, where "$f(x) \leq g(x)$" refers to the partial ordering in $\mathcal{T}$. In other words a complete lattice structure is imposed on the pixel intensity range. Usually $\mathcal{E}$ (the space of pixels) is taken to be either $\mathbb{R}^d$ (d-dimensional Euclidean space) or $\mathbb{Z}^d$ (d-dimensional discrete space), hence $\mathcal{F}$ corresponds respectively to the set of continuous or discrete images. Likewise various choices are available for $\mathcal{T}$, such as $\mathcal{T} = \overline{\mathbb{R}}^n$ ($\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$) and $\mathcal{T} = \overline{\mathbb{Z}}^n$. The case of $n > 1$ corresponds to the so-called multivalued images [85], while $n = 3$ includes specifically colour images. Namely, in the case of a multivalued image with n components, $\mathcal{T} = \mathcal{T}_1 \times \ldots \times \mathcal{T}_n$ is considered as the cartesian product of n complete lattices, and each mapping $f_i : \mathcal{E} \to \mathcal{T}_i, \;\; i \in \{1, \ldots, n\}$ is called a *channel* or *band* of the multivalued image.

Within this model, morphological operators are represented by mappings between complete lattices (i. e. the input and output images) with some additional properties such as increasingness and translation invariance. They are employed in combination with matching patterns, called *structuring elements* (SE), that are usually subsets of $\mathcal{E}$ (i. e. *flat* SE). Particularly, *erosion* and *dilation* constitute the fundamental blocks of MM, from the combinations of which several sophisticated operators can be derived. More precisely, given two complete lattices $\mathcal{L}$ and $\mathcal{M}$, from an algebraic point of view, an operator $\varepsilon : \mathcal{L} \to \mathcal{M}$ is called an erosion, if it is distributive over infima, i. e. $\varepsilon(\bigwedge_i P_i) = \bigwedge_i \varepsilon(P_i)$ for every collection $\{P_i\}$ of elements of $\mathcal{L}$. Similarly, $\delta : \mathcal{L} \to \mathcal{M}$ is called a dilation, if it is distributive over suprema, i. e. $\delta(\bigvee_i P_i) = \bigvee_i \delta(P_i)$ for every collection $\{P_i\}$ of elements of $\mathcal{L}$. As suggested in Ref. [240], dilation and erosion basically rely on three concepts: a ranking scheme, the extrema derived from this ranking and finally the possibility of admitting an infinity of operands. Yet, the first two are missing from multivalued images.

For example, if we apply the preceding notions to the case of continuous multidimensional grey-level images ($f : \mathbb{R}^d \to \overline{\mathbb{R}}$), it suffices to replace the partial ordering $\leq$ of $\mathcal{T}$ with the usual comparison operator in $\overline{\mathbb{R}}$, in order to induce a complete lattice structure on $\mathcal{T}$ and subsequently on $\mathcal{F}$ by means of Eq. (3.2.1), which will make the computation of extrema possible during erosion and dilation. Likewise, the inclusion operator "$\subseteq$" can be used with binary images ($f : \mathbb{R}^d \to \{0, 1\}$). However, if we now consider multivalued images ($f : \mathbb{R}^d \to \overline{\mathbb{R}}^n, \;\; n > 1$), it becomes problematic to find an ordering relation for the vectors of $\overline{\mathbb{R}}^n$, due to the fact that there is no universal method for ordering multivariate data.

In order to remedy this inconvenience, in the classic paper of Goutsias et al. [86] it is proposed to employ an adequate surjective mapping h to transform the image data into a more "suitable" space for morphological operators. More precisely, the idea of using a surjective mapping $h : \mathcal{T} \to \mathcal{L}$, where $\mathcal{T}$ is a non empty set and $\mathcal{L}$ a complete lattice, constitutes the theoretical support upon which several of the present multivariate morphological frameworks are based. Specifically, its importance lies in the fact that $\mathcal{T}$ is no longer required to be a complete lattice, since the ordering of $\mathcal{L}$ can be induced upon $\mathcal{T}$ by means of h :

$$\forall\, t, t' \in \mathcal{T}, \;\; t \leq_h t' \Leftrightarrow h(t) \leq h(t') \tag{3.2.2}$$

hence making it possible to construct h-morphological operators on $\mathcal{T}$. Consequently, one can deal with multivalued images, $f : \mathcal{E} \to \mathbb{R}^n$, through the use of a well chosen mapping $h : \mathbb{R}^n \to \mathcal{L}$, where $\mathcal{L}$ is a new, more suitable space for lattice based operations [85, 86].

Besides, given an adequate vector ranking scheme, the vector erosion ($\boldsymbol{\varepsilon}_b$) and dilation ($\boldsymbol{\delta}_b$) of a multivalued image $\mathbf{f}$ by a flat SE b, can be expressed immediately by means of the vector extrema operators $\sup_v$ and $\inf_v$ based on the given ordering:

$$\boldsymbol{\varepsilon}_b(\mathbf{f})(\mathbf{x}) = \inf_{v\,\mathbf{s}\in b} \{\mathbf{f}(\mathbf{x}+\mathbf{s})\} \tag{3.2.3}$$

$$\boldsymbol{\delta}_b(\mathbf{f})(\mathbf{x}) = \sup_{v\,\mathbf{s}\in b} \{\mathbf{f}(\mathbf{x}-\mathbf{s})\} \tag{3.2.4}$$

Therefore, the main obstacle preventing the extension of morphological operators to multivalued images, consists in defining an ordering relation that will induce a complete lattice structure on the set of vector pixel intensities.

3.2.3 Vector orderings

Especially in the last few decades, a lot of effort has been put in engineering a way of ordering vectors. Although numerous techniques for ordering multivariate data can be found in the literature [106, 158, 210, 260, 263, 264], according to the classical paper of Barnett (1976) [21], they can be classified into one of the following groups.

Marginal ordering (M-ordering) which corresponds to univariate orderings realised on every component of the given vectors:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n,\;\; \mathbf{v} \leq \mathbf{v}' \Leftrightarrow \forall\, i \in \{1, \ldots, n\}\,,\;\; v_i \leq v_i' \tag{3.2.5}$$

Data is ordered along each one of its channels independently from others, hence also the name *componentwise* ordering.

Conditional (sequential) ordering (C-ordering) in which vectors are ordered by means of some of their marginal components, selected sequentially according to different conditions. Whereas the components not participating in the comparison process are listed according to the position of their ranked counterparts. Hence, the ordering of the vectors is *conditioned* upon the particular marginal set of ranked components. Lexicographical ordering constitutes a widely known example of C-ordering employing potentially all the available components of the given vectors:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n,\;\; \mathbf{v} <_L \mathbf{v}' \Leftrightarrow \exists\, i \in \{1, \ldots, n\}\,,\;\; (\forall\, j < i,\; v_j = v_j') \;\wedge\; (v_i < v_i') \tag{3.2.6}$$

For vectors in a 3D space VWX, where we compare first dimension V, then W and finally X, we adopt the notation $V \rightarrow W \rightarrow X$. But of course one can also restrict the comparison process to use only a subset of the available components, as in Ref. [106]. C-orderings are most suitable for cases where one can establish a priority among the image channels.

Partial ordering (P-ordering), in which case "partial" is an abuse of terminology (Sec. 3.2.1), because not only there are total pre-orderings belonging to this class (Sec. 3.3.3), but algebraically partial orderings do not necessarily belong to this class either. That is why we will use the term P-ordering. P-orderings consist of approaches that partition the given vectors into equivalence classes with respect to order, rank or extremeness [261]. They are generally geometric in nature and account well for the inter-relations between components.

A simple example of P-ordering is the one based on the "peeling" of a multivariate sample (Fig. 3.1). First the convex hull of the entire sample is calculated. Consequently, the points on the border of the convex hull constitute the first class. The points belonging to

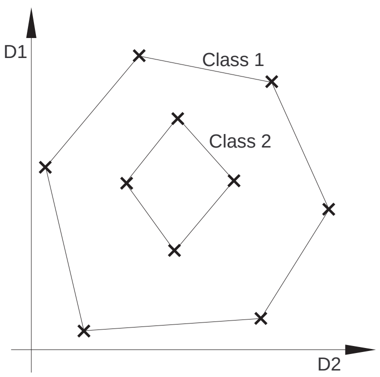


Figure 3.1: A P-ordering example in a bi-dimensional space $D1 \times D2$, based on the "peeling" principle.

the border of the convex hull of the interior constitute the second, and so forth. As a result, the entire sample becomes partitioned and ranked according to their class number, but of course no internal distinction is in place for the contents of the classes, hence it is a total pre-ordering.

Reduced ordering (R-ordering) in which vectors are first reduced to scalar values and then ranked according to their natural scalar order. A further categorisation of R-orderings consists in classifying them as distance orderings and projection orderings [166]. For instance, a R-ordering on $\mathbb{R}^n$ could consist in defining first a transformation $h : \mathbb{R}^n \to \mathbb{R}$, and then ordering the vectors of $\mathbb{R}^n$ with respect to the scalar order of their projection on $\mathbb{R}$ by h:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \;\; \mathbf{v} \leq \mathbf{v}' \Leftrightarrow h(\mathbf{v}) \leq h(\mathbf{v}') \tag{3.2.7}$$

According to the chosen transformation it is possible to obtain a total pre-ordering (h non-injective) or even a total ordering (h injective) [37]. An additional advantage of R-orderings lies in the fact that with an adequately chosen h, they can attribute equal priority to all components, unlike C-orderings.

As the aforementioned ordering groups are not mutually exclusive, their combinations as well as numerous implementational variants have led to several morphological frameworks with diverse properties. The next section will elaborate on the different approaches.

3.3 Approaches to multivariate MM

Despite the rich variety of multivariate morphological frameworks, there are in fact two main variables that are modified at each case, the extrema calculation method and the transformation (e. g. domain space change, etc) if any, that takes place on the image data before ranking; undoubtedly both influence the properties of the resulting operators.

In this section, we review in relative detail the approaches used for implementing multivariate morphological operators, with focus on colour morphology, primarily according to their vector ordering scheme, along with additional comments on their use with different preprocessing methods. However, instead of the conventional ordering categorisation presented in Sec. 3.2.3, it was decided to adopt the scheme employed by Chanussot in Ref. [37] where they are classified according to their algebraic properties, as given in Sec. 3.2.1. This

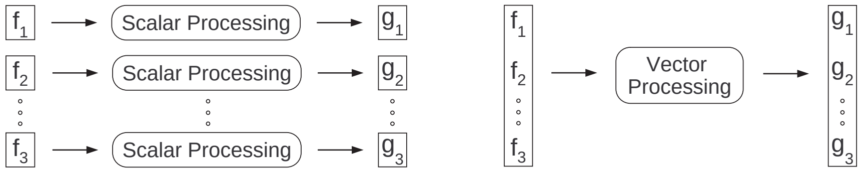


Figure 3.2: Marginal (left) and vector processing (right) strategies

choice was made with the aim of underlining the effect that these basic properties have on the end result of processing. Additionally, the vectors of the ordering relations that are mentioned in the sequel, are considered in $\mathbb{R}^3$, unless otherwise specified.

3.3.1 Processing strategies

Given a multivalued image, in practice there are two general methods of morphological processing: *marginal* (or *componentwise/scalar*) and *vector.*

Marginal Processing: It consists in processing separately each channel of the image. The inter-channel correlation is totally ignored, along with all information that could be potentially used in order to improve the quality of the result. Furthermore, the repetition of the processing procedure for each channel renders it expensive in terms of computational complexity. On the other hand, the marginal approach makes it possible to employ directly all methods offered by grey-level morphology (Fig. 3.2 left).

Vector Processing: This main alternative of the marginal approach, as its name implies, processes all available channels globally and simultaneously. Given that the vector pixels are considered as the new processing units, the correlation among the different channels is no longer ignored (Fig. 3.2 right). However, when compared to its marginal counterpart, the most important inconvenience of the vector approach appears to be primarily the need for adapting the existing algorithms in order to accommodate vector data; thus leading often to slower implementations than their scalar versions. The rest of this section elaborates on each approach as well as on their variants.

3.3.2 Partial ordering based approach (Marginal)

The so-called *marginal* processing strategy, despite being presented usually as an alternative to *vector*, is as a matter of fact no more than just its variant, as it employs the partial ordering defined in Eq. (3.2.5). Obviously, there can be vectors that may not be comparable under this ordering relation, for instance $\mathbf{a} = [7, 2]^T$ and $\mathbf{b} = [3, 4]^T$. Nevertheless, this does not prevent the definition of valid morphological operators based on extrema computed by means of this ordering [242]. Furthermore, it makes it possible to employ all tools offered by grey-level morphology with no need for special adaptation steps. For instance, the erosion and dilation expressions given in Sec. 3.2.2 become equivalent to:

$$\boldsymbol{\varepsilon}_b(\mathbf{f})(\mathbf{x}) = [\varepsilon_b(f_1)(\mathbf{x}), \ldots, \varepsilon_b(f_n)(\mathbf{x})]^T \tag{3.3.1}$$

$$\boldsymbol{\delta}_b(\mathbf{f})(\mathbf{x}) = [\delta_b(f_1)(\mathbf{x}), \ldots, \delta_b(f_n)(\mathbf{x})]^T \tag{3.3.2}$$

where ε_b and δ_b denote respectively the scalar erosion and dilation operators with a SE b. For instance, Gu in Ref. [89] has employed the marginal approach for establishing multivalued

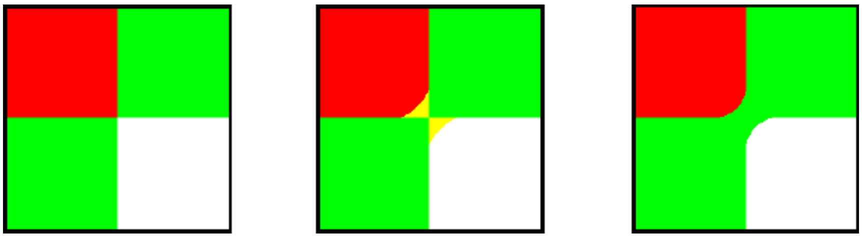


Figure 3.3: Example of false colours. From left to right, the original image, the result of a marginal median filter and of a vector (lexicographical) median filter.

morphological operators applied to moving object segmentation and tracking. Another example is given by Aptoula et al. in Ref. [14], where marginal morphological operators are employed with the purpose of galaxy detection from multispectral data. Moreover, a generalisation of the marginal approach was given by the matrix morphology theory of Wilson [293], which was based on the work of Heijmans and Ronse [112].

Despite its implementational simplicity, marginal ordering suffers mainly from two disadvantages: not accounting for inter-channel information as well as the risk of altering the spectral composition of its input. More precisely, as each component is processed independently, any eventual correlation among them is totally ignored, hence rendering this approach unsuitable for images with highly correlated components (e. g. RGB colour images) [17]. A possible solution to this problem, as proposed in Ref. [86], consists in applying a decorrelating transformation (e. g. maximum noise fraction transform (MNF), principal component analysis (PCA), discrete cosine transform, a colour space with independent components, etc) prior to ordering. Nevertheless, these decorrelating transformations also introduce an additional computational burden.

Furthermore, there is absolutely no guarantee that marginally processed vectors belong to the input image. The lack of vector preservation constitutes an undesirable effect for several applications. For example, in the case of colour image processing this would lead to the appearance of new hues (also known as *false colours*) and thus deteriorate the visual quality of the result, and in particular the colour balance and the object boundaries; whereas the effect near the spatial edges, leads to the so-called *edge jitter*. This effect is illustrated in Fig. 3.3, where the application of a marginal median filter leads to the appearance of yellow, as the result of additive combination of red and green.

A detailed study of the problem of vector preservation can be found in Refs. [126, 258]. According to the last reference, the only way to use morphological operators without generating new vectors is to impose an order on the vector space by means of an (pre-)ordering verifying the totality constraint; besides, this is why the majority of the published articles on multivalued morphology deal with total (pre-)orderings. Yet, vector preserving approaches are in the same time limited by this property, due to the restriction imposed on their output (i. e. the output must be a vector from the input set); whereas the marginal approach has access to a much broader range of output values, an advantage which becomes most valuable during noise reduction [51].

Further attempts motivated by the intuitiveness of the marginal approach have resulted in improvements. Namely, Serra in [242] presents an intermediate form between a M-ordering and a C-ordering, while an application of marginal ordering in the HSV colour space is introduced by Weber and Acton in Ref. [287], where the aforementioned undesirable effects are reduced by shifting the hue origin. Moreover, Al-Otum in Ref. [1], proposes the *corrected componentwise morphological algorithm* with the goal of preventing the appearance of new vectors. It consists in replacing each new vector of the output with its closest vector from the original image, chosen with the help of a Mahalanobis distance based error function.

In brief, either accompanied by additional transformations or not, the marginal processing strategy uses conventional grey-level operators and pixels are still treated as scalar values, whereas its potential with colour data is severely limited by its lack of vector (or colour) preservation.

3.3.3 Total pre-ordering based approach

Contrary to the marginal approach, by means of the additional property of totality, all vectors become comparable and as a result it is possible to construct a totally ordered lattice structure. Hence pixels can be manipulated as whole vectors, and consequently the risk of introducing new vectors is completely eliminated [258].

All pre-orderings however share a common drawback, which is the relaxation of the anti-symmetry constraint. Thus distinct vectors can eventually end up being equivalent. That is why additional measures become necessary, in order to resolve the ambiguity of eventually multiple extrema. For instance an useful solution, though partial, proposed by Comer and Delp [51] for this problem, consists in selecting the output vector according to its position in the SE. Besides, the combination of an adequately chosen pre-ordering with an ordering for tie-breaking purposes has also been successfully used in practice [9].

Reduced total pre-orderings: Total pre-orderings can be obtained with R-orderings employing a non-injective reduction transformation [37]. Distance measures are typical examples of such transformations; in fact they account for the majority of the proposed R-orderings. Colour distances however do not necessarily represent the corresponding perceptual differences, unless they are chosen in combination with a perceptually linear colour space. That is why distance based colour orderings may lead to unexpected results, if used with non perceptually uniform spaces.

This first variant ranks vectors according to their distance from a reference vector $\mathbf{v}_{ref}$:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow d(\mathbf{v}, \mathbf{v}_{ref}) \leq d(\mathbf{v}', \mathbf{v}_{ref}) \tag{3.3.3}$$

where $d(\cdot, \cdot)$ represents a distance measure. This particular case of distance based colour ordering has been studied by Angulo [9] and Sartor and Weeks [235], where the pitfall of multiple extrema is avoided by combining the distance computation with the lexicographical cascade, Eq. (3.2.6):

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow [d(\mathbf{v}, \mathbf{v}_{ref}), v_1, \ldots, v_n]^T \leq_L \left[d(\mathbf{v}', \mathbf{v}_{ref}), v'_1, \ldots, v'_n\right]^T \tag{3.3.4}$$

In case the reference vector is the origin, Eq. (3.3.3) becomes equivalent to using the norms of the vectors. A variant of Eq. (3.3.3), proposed for the morphological processing of multispectral remote sensing data [211], which eliminates the need for a reference vector, consists in associating each vector with the sum of its distances from the other vectors of a family $\{\mathbf{v}_j\}$:

$$\forall\, \mathbf{v}_k, \mathbf{v}_l \in \{\mathbf{v}_j\}\,, \;\; \mathbf{v}_k \leq \mathbf{v}_l \Leftrightarrow \sum_j d(\mathbf{v}_k, \mathbf{v}_j) \leq \sum_j d(\mathbf{v}_l, \mathbf{v}_j) \tag{3.3.5}$$

This additional advantage however comes at an elevated computational cost. Furthermore, although this approach associates each vector with a scalar value denoting its "spectral purity", as far as colour images are concerned, the infimum of a family of vectors calculated in this way corresponds to the notion of "median vector" and consequently does not carry the significance of a "minimum" in the numerical sense. There is also another property worth mentioning [266], concerning Eq. (3.3.5) and implementations of Eq. (3.3.3), if the reference vector is chosen to be "in the middle" (e. g. the median or the average). It concerns the instability of the supremum computed by means of these orderings. Despite the remarkable stability of the infimum under the same conditions, even a slightly varied input can radically change the least upper bound. Consequently, the dilation operator given in Eq. (3.2.4), that makes use of the supremum becomes unsound from a practical point of view.

The choice of the distance measure is of course another key topic. Theoretically, any kind of metric can be used, and so it is in practice [68, 211]. In Refs. [1, 268] for instance, Eq. (3.3.3) is used coupled with the Mahalanobis distance for ordering the components of colour images in different colour spaces, with "black" being the reference vector. While the same distance is employed along with Eq. (3.3.4) with the LSH colour space in Ref. [9]. A further example of Euclidean norm based colour ordering can be found in Ref. [51], as applied in the RGB colour space.

Moreover, an original approach implemented in a phenomenal colour space appears in Ref. [2], where a saturation based combination of the hue and intensity planes is proposed. Specifically, the vector pixels contained within the SE are reduced into scalars by means of a weighted combination of normalised vector angle (d_a) and Euclidean distances (d_E) from the vector $\mathbf{v}_c$ at the centre of the SE:

$$h(\mathbf{v}) = w(\mathbf{v}, \mathbf{v}_c) \cdot d_a(\mathbf{v}, \mathbf{v}_c) + (1 - w(\mathbf{v}, \mathbf{v}_c)) \cdot d_E(\mathbf{v}, \mathbf{v}_c) \tag{3.3.6}$$

where $w(\cdot, \cdot) \in [0, 1]$ denotes the weight computed based on the saturation levels of the given vector pixels. The underlying idea is to favour hue differences (d_a) when saturation levels are high, while intensity differences (d_E) gain more importance with achromatic pixels.

Of course the non-injective transformation choices are by no means limited by these few variants. For instance, in Ref. [142], colour vectors are ordered based on the number of times they appear in the input image. Nevertheless, distance based R-orderings hold the potential of accounting for all dimensions without privileging any of them, a property which becomes particularly useful in case there is no predefined order of importance among the available channels (e. g. RGB colour images). Otherwise any transformation capable of realising the necessary reduction may be used. For example in Ref. [51], an additional R-ordering is presented, applied to RGB colour images, where vectors are reduced to scalars with the help of a linear weighted combination:

$$h(\mathbf{v}) = w_1 \cdot v_1 + \ldots + w_n \cdot v_n, \tag{3.3.7}$$

hence making possible the arbitrary prioritisation of colour channels by means of the $\{w_i\}$ coefficients. Finally, a more complicated ordering scheme takes place in Ref. [150], where a new R-ordering based on ordinal first principal component analysis is introduced and the derived operators are applied to edge detection on RGB colour images.

Conditional total pre-orderings: As previously mentioned in Sec. 3.2.3 C-orderings restrict the ordering process to only one or more components of the given vectors, while the others are conditioned upon them. That is why C-orderings are suitable for situations where certain channels are more "privileged" than others. Besides, unless all vector components participate in the ordering process, the resulting C-ordering is bound to be a total pre-ordering, thus sharing their aforementioned inconveniences. For example, in the case where only the first component is employed [106]:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow v_1 \leq v_1' \tag{3.3.8}$$

Hence, the two distinct vectors $\mathbf{a} = [7, 2]^T$ and $\mathbf{b} = [7, 3]^T$ would be considered equivalent according to Eq. (3.3.8). The main problem of C-orderings concerns of course the choice of the ordered components. Obviously, ordering vectors along only some of their components is practically justifiable only if the given components represent sufficiently the vectors.

In the general case, this approach attributes far too much significance to the selected components, disregarding all others. While usually this constitutes a severe problem, there are cases where it becomes invaluable, For instance when it is a priori known that certain channels are noise free. In fact, C-orderings are almost always used in combination with a suitable domain space change, so that the data of interest, in its majority, will lie in only some of the channels.

An example can be found in Ref. [273], where a C-ordering on the HSV colour space is employed, ignoring the hue component, hence avoiding the problem of false colours, despite the use of a non-total ordering. Specifically, given two HSV triplets $\mathbf{v} = (h, s, v)$ and $\mathbf{v}' = (h', s', v')$:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow \begin{cases} v \leq v', & \text{or} \\ v = v' & \text{and} \;\; s \geq s' \end{cases} \tag{3.3.9}$$

In fact, this also constitutes a fine example of a combined ordering; more precisely, the result can also be considered as a P-ordering where the ordered groups contain the colours of equal value and saturation with no internal distinction whatsoever as hue is not taken into account. An application of the same ordering on colour image skeletonisation can be found in Ref. [5].

3.3.4 Total ordering based approach

Total orderings, from a theoretical point of view, have two main advantages that render them more suitable for vector ordering, as far as multivariate MM is concerned. First, thanks to their totality, they are vector preserving, and contrary to pre-orderings, as they verify the anti-symmetry constraint the computed extrema are unique. That is why the majority of the attempts concentrated on extending morphological operators to multivalued images are based on total orderings. In particular, the lexicographical ordering (C-ordering) along with its variants is among the most implemented choices in colour morphology.

However, the uniqueness of extrema takes a serious toll, since the prioritisation of certain vector components becomes inevitable [37]. That is why they are almost always used in combination with a suitable domain transformation (e. g. PCA, a colour space with independent components, etc) that will place the "interesting" part of the data in the first few channels. Nevertheless, as it will be subsequently presented some implementations tend to be more "symmetric" than others.

Lexicographical ordering: Lexicographical ordering, introduced in Eq. (3.2.6), is undoubtedly the most widely employed total ordering within this context. As a conditional ordering, it is most suitable to situations where an order of "importance" exists on the

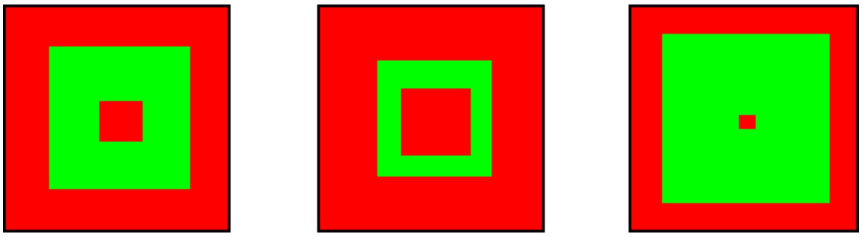


Figure 3.4: original image (left), results of applying a vector dilation based on a lexicographical ordering (RGB - middle) (GRB - right) with a 21×21 square SE

available channels, either inherently or artificially created by means of an appropriate transformation. This prioritisation of certain vector components with respect to others, constitutes also the main drawback of lexicographical ordering. To explain, Fig. 3.4 provides an example of the priority attributed to the first component during lexicographical ordering. More precisely, a vector dilation is applied on a RGB colour image (Fig. 3.4, left) and as red is the head component, it dominates visibly over green (Fig. 3.4, middle). Whereas if we permute the channels as GRB, the effect is reversed in favour of green (Fig. 3.4, right).

As far as colour MM is concerned, the inherent separation of chromatic and achromatic information in colour spaces such as $L^*a^*b^*$, HSV, and LSH, makes lexicographical ordering a prime ordering choice, since according to Sec. 2.2, achromatic information alone, is often sufficient for the recognition of most objects. That is why, unless the application demands it, most colour MM implementations are based on brightness first lexicographical configurations. A first example on the HSV colour space can be found in [157]:

$$\forall\, \mathbf{v} = (h, s, v), \mathbf{v}' = (h', s', v'), \;\; \mathbf{v} < \mathbf{v}' \Leftrightarrow \begin{cases} v < v' \;\text{ or} \\ v = v' \;\text{ and }\; s > s' \;\text{ or} \\ v = v', s = s' \;\text{ and }\; h < h' \end{cases} \tag{3.3.10}$$

which however does not take into consideration the 2π periodicity of the hue component [103, 207]. This same ordering is used in Refs. [155, 272] for respectively colour granulometries and median filtering. Of course there can also be specific situations where chromatic information is more significant, like for instance in Ref. [196], hence hue can be compared first in the lexicographical cascade.

Other applications of lexicographical ordering in the HSI space include Ref. [197], where reconstruction based morphological operators are developed with the goal of brightness elimination, using an intensity based ordering scheme. The polar representations of the almost perceptually uniform $L^*u^*v^*$ [207] and $L^*a^*b^*$ [104] spaces have also been used together with a lexicographical ordering. The same principle is additionally employed in Ref. [199] with the purpose of noise elimination. Moreover, a thorough study of the potential of this ordering in the HLS space is provided in Ref. [102], while Ref. [198] shows the results of a comparative application of this ordering in different hue based colour spaces. Besides, following the variety and number of lexicographical approaches developed for phenomenal spaces, Angulo [8] proposed a “unified framework” consisting of using the lexicographical ordering in the LSH colour space.

As the majority of lexicographical comparisons are determined by the first components [104], variants of the classical lexicographical ordering have been proposed, with the purpose of better tuning the priority as well as degree of influence of each component. A first group of variants is based on preceding the lexicographical cascade by a component representative of the entire vector. This principle is used for example along with an Euclidean norm in Refs. [223, 224]:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow [\|\mathbf{v}\|, v_1, \ldots, v_n]^T \leq_L [\|\mathbf{v}'\|, v'_1, \ldots, v'_n]^T \tag{3.3.11}$$

as well as in Refs. [9, 235], that employ a distance with respect to a reference vector, Eq. (3.3.4). Of course, there is no limit to the number or type of functions that can be used according to this principle. Other examples include the use of the maximum and minimum of the compared components in the case of RGB colour images, as well as their weighted combinations [11]. Another type of extension to the classical lexicographical ordering consists in using a user defined parameter α, in such a way that it can modify the degree of influence of the first component. For instance, the *α-modulus lexicographical ordering*, introduced by Angulo [6, 8]:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n,\;\; \mathbf{v} \leq \mathbf{v}' \Leftrightarrow [\lceil v_1/\alpha \rceil, v_2, \ldots, v_n]^T \leq_L [\lceil v'_1/\alpha \rceil, v'_2, \ldots, v'_n]^T \tag{3.3.12}$$

was implemented on LSH. Specifically, given integer valued colour images, by dividing the first component with a parameter α, and then rounding it off to the next closest integer, a sub-quantisation of the first vector dimension is realised, hence forming larger equality groups within this dimension. Consequently, a higher number of comparisons is expected to reach the second dimension. This approach is further detailed in the next chapter.

Bit mixing based ordering: Bit interlacing (or mixing) constitutes an innovative R-ordering, originally developed for RGB colour images, aiming mainly to eliminate the unavoidable asymmetry which results from the application of total orderings. Specifically it employs an injective transformation exploiting the binary representation of each component in order to impose a total order on the vector space [37, 38, 40].

Given a vector $\mathbf{v}$, with each component coded in k bits, the corresponding reduction transformation $h : \mathbb{Z}^n \to \mathbb{Z}$ is formulated as:

$$h(\mathbf{v}) = \sum_{m=1}^{k} \left\{ 2^{n\cdot(k-m)} \cdot \sum_{i=1}^{n} 2^{n-i} \cdot v_{i,m} \right\} \tag{3.3.13}$$

where $v_{i,m}$ denotes the m^{th} bit of the i^{th} component of $\mathbf{v}$. Hence, the resulting binary representation of $h(\mathbf{v})$ becomes:

$$v_{1,1}v_{2,1}\cdots v_{n,1}v_{1,2}v_{2,2}\cdots v_{n,2} \;\;\cdots\;\; v_{1,k}v_{2,k}\cdots v_{n,k} \tag{3.3.14}$$

Besides being endowed with all the qualities of a total ordering, bit mixing provides a more symmetrical approach than its lexicographical counterpart as dimensions are mixed in bit level. Of course some dimensions continue to be more privileged than others, with the degree of that privilege being proportional to the significance of the bit position that they occupy. Furthermore, a finer grained symmetry can be obtained by modifying the original mix order [37]. Then again, there are several situations where data channels need to be processed with a certain priority. Bit mixing can easily respond to this requirement by placing the important vector components to more significant bit positions. Additionally, bit interlacing can also be used as a means of fusing the channels, hence allowing for instance the use of the watershed transformation on colour images [39].

From a theoretical point of view, this approach aims to fill a given multi-dimensional space using a "balanced" *space filling curve* (SFC) with respect to the available dimensions. In the case of total orderings, these curves (e. g. Peano curve) pass through all vector coordinates of the space under consideration, hence vectors can be ordered according to their position on it. On the other hand, according to Ref. [253] the main inconvenience of a total ordering obtained in this way is its lack of physical interpretation.

3.3.5 Other approaches

Besides the previously presented ordering methodologies, there have been also some rather "unconventional" orderings or extremum calculation approaches, in the sense that they either do not consist in simply using a standard vector ordering scheme, or they are developed for a particular form of image data, or even combine additional theories with the goal of achieving an efficient solution.

For example, following the success of fuzzy morphology with grey-level images [27, 57], a couple of attempts to apply the fuzziness concepts to multivariate images have already been carried out. Köppen et al. in Ref. [135], introduced *fuzzy Pareto morphology*, a means of computing multivariate extrema based on the notion of Pareto sets, a concept belonging to the field of multi-criteria optimisation combined with fuzzy subsethood. However, since there is no underlying binary ordering relation, it results in pseudo-morphological operators. A second attempt was made by Louverdis et al. [156], who employ a direct fuzzification of the vector pixels in the HSV colour space, that are subsequently ranked according to a lexicographical ordering scheme. The morphological operators based on this total pre-ordering have been compared in the original article with their counterparts in Ref. [157] resulting from the use of a similar lexicographic ordering also in the HSV space. According to their experiments, the fuzzy version has slightly better noise reduction capabilities as well as less sensitivity to distortions.

Some further original approaches aiming to compute multivariate extrema, include the one proposed in Ref. [289], developed with the purpose of colour object detection in the HLS space using vector projection measures in combination with vector SE, as well as the graph based methodology introduced by Lezoray et al. [147], employing the minimum spanning tree algorithm among the pixels under the SE. An ordering based on labels has been additionally proposed by Ronse and Agnus [228], while Ronse mentions an ordering using bundle lattices for polar colour representations in Ref. [227]. Moreover, Zaharescu et al. have explored in Ref. [298] the potential of the triangle representation of colours, which leads to a R-ordering, while Mojsilovic and Soljanin [179] have employed a method based on a quantisation using Fibonacci lattices, in order to obtain a partial ordering. And finally Gibson et al. [81] have relied on local convex hull computations for locating multivariate extrema, whereas Busch and Eberle [31] have proposed a conditional ordering based on semantic principles.

3.3.6 Synopsis

This section presented the different approaches that have appeared so far in the literature, with the goal of extending morphological operators to multivalued images. A summary of the related references is given in Table 3.1. As far as colour morphology is concerned, although the ordering methodologies in use vary considerably, the colour space choices appear to be largely in favour of phenomenal spaces. On the other hand, the variety of vector orderings render

Data space	Properties		
	¬ VP	VP ∧ PB	VP ∧ ¬ PB
Generic	[51, 86, 89, 179]	[7, 31, 38, 40, 81, 179, 228]	[1, 7, 51, 68, 135] [142, 147, 150, 211, 255, 298]
HSV	[287]	[5, 97, 155, 156, 157, 196, 272, 273]	
L*a*b*		[104]	
L*u*v*		[207]	
HLS/HSI/LSH		[9, 8, 11, 102, 197, 199, 200, 289]	
C-Y		[2, 235]	
Complex		[223, 224]	[289]

Table 3.1: Synoptic table of references developing multivariate morphological operators, organised according to the properties of the employed ordering scheme and its corresponding implementation space (VP: vector preserving, PB: prioritisation of bands)

it necessary to further proceed to a comparative study that will help refine their properties, prior to making a definitive choice.

3.4 Experimental comparison of selected vector orderings

Having provided an insight into several morphological frameworks, either developed specifically for colour images or able to accommodate them, in this section we carry out a brief series of comparative tests with two image processing tasks, with the aim of better determining their properties and relative performances. Previous comparative studies concerning colour morphology may be found in Refs. [51, 61, 198, 279, 280], where at most three orderings are compared in a single application. They can be hardly considered extensive, since there is an overwhelming number of crucial variables that need to be taken into account in order to achieve a fully objective performance measure (e. g. number of images, type of images, application types and parameters, correlation of channels, colour spaces, etc). That is why the aim of this section is *not* to realise a benchmark of the different approaches, but rather to provide indications on the suitability of each ordering for particular input properties, as well as to assert the remarks made in the previous sections with experimental results.

Method	Equation	Acronym	Configuration
Marginal	(3.2.5)	M	-
Lexicographical	(3.2.6)	L	-
Reference vector based	(3.3.3)	R	median as reference
Reference vector based	(3.3.3)	N	origin as reference
Cumulative distance based	(3.3.5)	C	Euclidean distance
Reduced lexicographical	(3.3.11)	rL	L_2 norm
α-modulus lexicographical	(3.3.12)	αL	$\alpha = 10$

Table 3.2: The orderings participating in experimental comparison

For the sake of simplicity only certain approaches have been considered during the tests, chosen subjectively as representative members of the classification given in Sec. 3.2.1. Their names and configurations are given in Table 3.2. As far as their implementation is concerned, vector comparisons were hard coded without the use of any indexes [258], within a generic

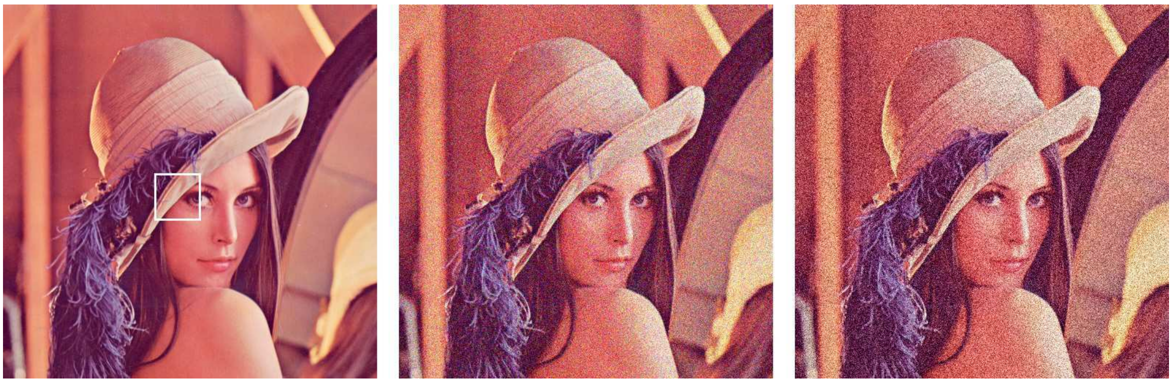


Figure 3.5: From left to right: the original image "Lenna", corrupted with uncorrelated Gaussian noise ($\sigma = 32$, $\rho = 0.0$) and corrupted with correlated Gaussian noise ($\sigma = 32$, $\rho = 0.9$). The white square represents the enlarged results in figure 3.6

Java based framework. As to the applications under consideration, it has been chosen to use noise reduction and texture classification, for both of which effective means of quantitative performance assessment are available.

3.4.1 Noise reduction

During this first test, the relative performances of the vector ordering schemes are examined in terms of noise reduction quality. As a quantitative measure, the *normalised mean squared error* (NMSE) is used:

$$\text{NMSE} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left\|\mathbf{f}(i,j) - \mathbf{f}'(i,j)\right\|^2}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left\|\mathbf{f}(i,j)\right\|^2} \tag{3.4.1}$$

where N and M represent the image dimensions while $\mathbf{f}(i,j)$ and $\mathbf{f}'(i,j)$ denote respectively the vector pixels at position (i,j) for the original and filtered images. The tests have been repeated with a number of RGB colour images of various content, however here we exhibit the results obtained for the "Lenna" image (Fig. 3.5). The subjects, all of size 512×512 pixels and 24 bits per pixel, processed with integer precision in $[0,255]^3$, are first contaminated with zero-mean additive Gaussian noise, $\sigma = 32$ and correlation factor $\rho = 0.0$ (Table 3.3) and $\rho = 0.9$ (Table 3.4). Other noise distributions that have also been tested, include double-exponential and uniform, however as the influence of the type of noise distribution on the resulting performances was observed to be minimal, they were omitted from tables 3.3 and 3.4 for the sake of clarity.

The filter employed for smoothing is open-close close-open (OCCO). More precisely, let first $\boldsymbol{\gamma}_b$ and $\boldsymbol{\phi}_b$ denote respectively the vector opening and closing operators:

$$\boldsymbol{\gamma}_b(\mathbf{f}) = \boldsymbol{\delta}_b(\boldsymbol{\varepsilon}_b(\mathbf{f})) \tag{3.4.2}$$

$$\boldsymbol{\phi}_b(\mathbf{f}) = \boldsymbol{\varepsilon}_b(\boldsymbol{\delta}_b(\mathbf{f})) \tag{3.4.3}$$

In which case, OCCO is defined as the pixelwise average of open-close and close-open:

$$\text{OCCO}_b(\mathbf{f}) = \frac{1}{2}\boldsymbol{\gamma}_b(\boldsymbol{\phi}_b(\mathbf{f})) + \frac{1}{2}\boldsymbol{\phi}_b(\boldsymbol{\gamma}_b(\mathbf{f})) \tag{3.4.4}$$

Colour	Vector Orderings						
spaces	M	L	N	αL	C	R	rL
RGB	**0.78**	2.29	2.10	2.21	3.10	3.07	2.10
GBR	**0.78**	2.33	2.10	2.19	3.10	3.07	2.10
L*a*b*	**0.92**	2.23	2.50	2.05	3.13	3.02	2.50
VSH	**0.80**	2.16	2.32	2.16	2.56	2.54	2.32
LSH	**0.79**	2.14	2.29	2.14	2.53	2.46	2.21

Table 3.3: NMSE×100 values for the "Lenna" image obtained against uncorrelated Gaussian noise

given the size of the input, the SE b is chosen as a cross of size 3×3. Larger sizes have been also tested but the relative performances remained almost the same. The OCCO filter is chosen primarily due to its effective combination of the basic morphological operators as well as for its suitability against the chosen noise type. On the other hand, taking the average of the intermediate results, obviously prevents the preservation of the initial vectors. Nevertheless, this is a desired property in the context of noise reduction as it enables a better approximation to original noise-free values. Different colour spaces have been also employed during smoothing for mainly two reasons; first in order to simulate various correlation levels among the channels as well as for testing the effect that uneven distributions of the intensity information among the channels have on the underlying ordering schemes. Specifically, it was chosen to use the following spaces: RGB, L*a*b* with the D65 white point Eq. (2.5.3), cylindrical HSV Eq. (2.7.1) (permuted as VSH) and LSH based on the L_1 norm Eq. (2.7.5). As far as the processing of the 2π periodic hue component is concerned, all hue values were replaced by their distance to a reference hue [103, 207] prior to filtering:

$$h \in [0, 2\pi], \quad h \div h_{ref} = \begin{cases} |h - h_{ref}| & \text{if} \quad |h - h_{ref}| < \pi \\ 2\pi - |h - h_{ref}| & \text{otherwise} \end{cases} \tag{3.4.5}$$

The reference hue has been set as $h_{ref} = 0$, i. e. red, and the following ordering has been subsequently used:

$$\forall\, h, h' \in [0, 1], \quad h < h' \Leftrightarrow h' \div h_0 < h \div h_0 \tag{3.4.6}$$

Table 3.3 shows the NMSE results of the orderings within five colour spaces against uncorrelated Gaussian noise. The entries corresponding to the best performance of their row (colour space) are in bold. According to the obtained values, the overall superiority of the marginal approach over its vector counterparts is remarkable.

As stated earlier the reduced smoothing capacity of vector approaches is part of the trade-off between noise reduction capability and vector preservation. In brief, even with a maximised processing symmetry among the available channels the final result of vector openings and closings in Eq. (3.4.4) will necessarily be one of the input vectors. And as such it is natural for the marginal strategy to outperform, by having access to a much broader range of output values, that are not necessarily included however in the original image.

A second remark concerns the relatively high error rates of the total orderings in the highly correlated RGB space, which is simply due to the fact that L and αL inevitably prioritise the vector components. Consequently, the green and especially blue channels influence the outcome of vector comparisons much less than red, thus resulting in poor smoothing quality in the last two channels. The effect of prioritisation is particularly visible with lexicographical ordering. An exception to this remark is rL, which although total, results always in a NMSE value almost identical to norm based reduced ordering (N), due to the vector norm that occupies the first position, as indicated in Eq. (3.3.11), during the lexicographical comparison of vectors. Therefore, as the vast majority of comparisons is decided by means of the first vector component, from a practical point view it appears to behave exactly as a norm based R-ordering.

As far as the three distance based approaches are concerned (R, C, N), the norm based ordering takes the lead while its distance based counterparts provide quite unsatisfactory results. Thanks to the non-injective scalarisation that takes place, no distinction is made among the channels during comparison. This last important property becomes more evident within the GBR space, a permutation of RGB, where they achieve the exact same performances thus underlining their robustness against situations where it is not a priori known in which channel takes place the majority of the "important" information or even which channel is the most corrupted.

Furthermore, the $L^*a^*b^*$, VHS and LSH colour spaces cause significant changes in the results as they are far less correlated than RGB and the brightness information is moved exclusively to the first channel. As expected, the αL and L improve their smoothing rates substantially, since their prioritisation works this time in their favour by rendering the brightness dimension more "exploitable". It is also presumed that the finer grained symmetry of the vector components during comparison is what makes it possible for the α-modulus lexicographical ordering to surpass its standard counterpart. Nevertheless, the marginal ordering still provides the best results.

On the other hand, Table 3.4 shows the NMSE results obtained against highly correlated Gaussian noise. Although the relative performances of vector approaches have stayed the same, this time they are clearly much closer to that of marginal processing.

Colour	Vector Orderings						
spaces	M	L	N	αL	C	R	rL
RGB	**0.79**	0.97	0.87	0.97	1.35	1.34	0.87
$L^*a^*b^*$	**0.85**	0.92	1.34	0.88	1.43	1.38	1.34
VSH	**0.79**	0.94	1.09	0.92	1.34	1.56	1.09
LSH	**0.79**	0.89	0.95	0.87	1.26	1.51	0.95

Table 3.4: NMSE×100 values for the "Lenna" image obtained against highly correlated Gaussian noise

M	L	N	αL	C	R	rL
1.36	1.00	1.49	1.01	2.03	2.22	1.54

Table 3.5: Relative durations for the application of the OCCO filter on the Lenna image with a 3×3 cross shaped SE with respect to the most efficient (L).

As far as the computation times are concerned, Table 3.5 illustrates the relative duration of a single application of the OCCO filter for each ordering. Apparently L and αL are the

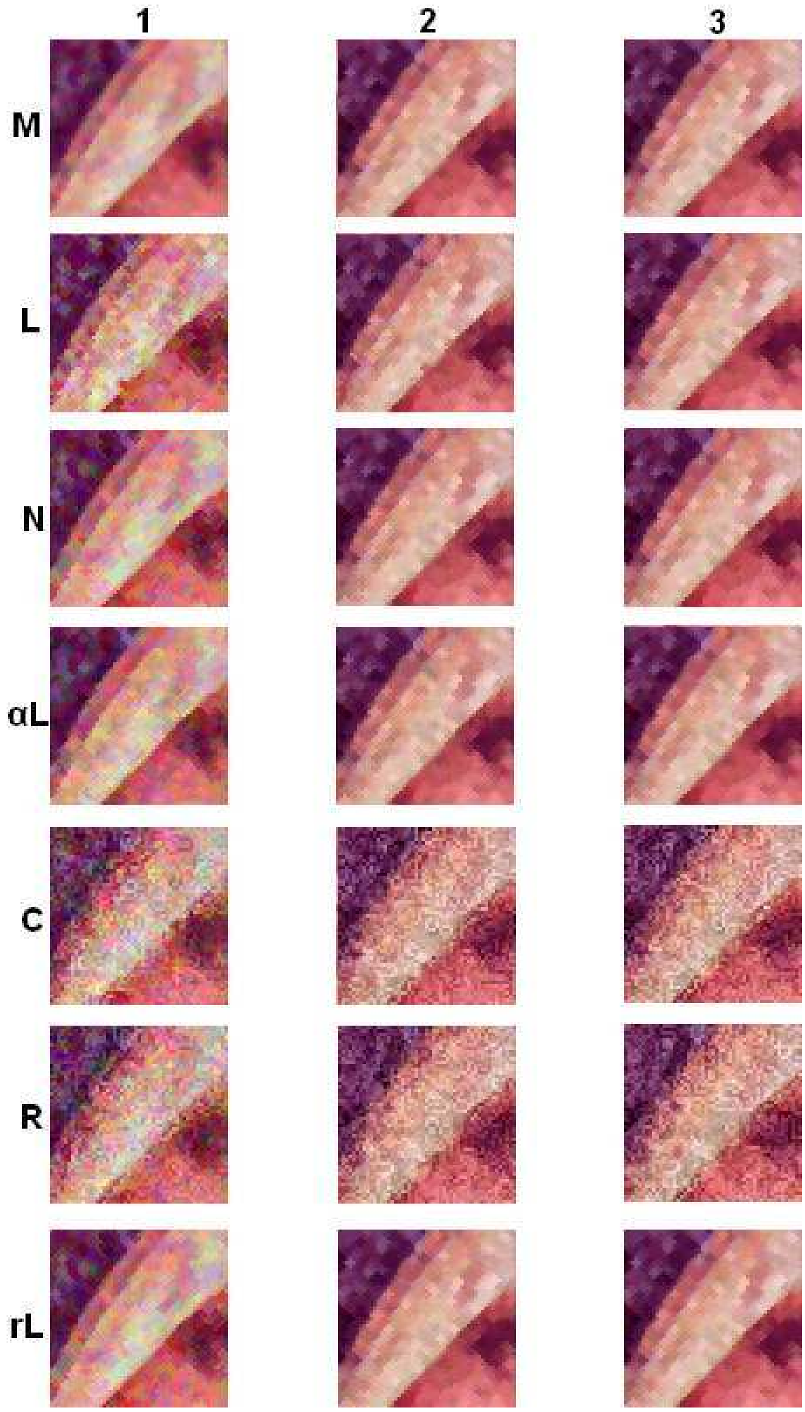


Figure 3.6: Enlarged results of the area contained within the white square of Fig. 3.5, obtained on the RGB colour space against uncorrelated Gaussian noise (column 1), on the RGB colour space against correlated Gaussian noise (column 2), and on the LSH colour space against correlated Gaussian noise (column 3)

Figure 3.7: Examples of the 68 textures of Outex 13 [162].

fastest among them, closely followed by the marginal approach, while the R-orderings suffer from the extensive distance calculations.

In conclusion, the main points of this discussion include the overall superiority of the marginal approach, particularly against uncorrelated noise. On the other hand both marginal and vector approaches exhibit similar performances in the case of correlated noise (Fig. 3.6). Although vector approaches tend to be more complex from an implementational point of view, they offer the possibility of smoothing without the risk of introducing new vectors. Additionally, the results have also underlined the importance of the colour space, as it highly influences the performance of the employed ordering.

3.4.2 Texture classification

As far as texture classification is concerned, here we employ the colour textures of Outex13 (Fig. 3.7) [191]. This set contains 68 different textures, with 20 images from each obtained with 100 dpi resolution under an incandescent CIE A light source. The total amount of 1360 images have been evenly divided as training and test sets.

The question whether colour should be processed separately or jointly from texture is still an open problem [162], and vector morphological feature extraction operators represent in their majority the latter case [101]. As colour texture descriptor we employ the morphological version of the autocorrelation operator, namely morphological covariance, which is detailed in Sec. 6.3. In short, it consists in describing its input through the variation of volume (i. e. sum of pixel values), that results from successive erosions by a pair of points separated by a vector of various lengths and orientations. Thus it provides information on the periodicity as well as anisotropy of textures.

Since the SE contains only 2 pixels, the use of R and C approaches is of no practical interest. The covariance based feature vectors have been calculated using four directions for the point pairs (0°, 45°, 90°, 135°), each along with distances ranging from 1 to 49 pixels in steps of size two. Consequently 25 values are available for each direction, making a total of 100 values for every image channel after concatenation. Furthermore, as far as the conversion from RGB to $L^*a^*b^*$ is concerned, the proper transformation matrix calibrated to the CIE A white point of the acquisition apparatus has been used [162]. The classification process is

Colour	Vector Orderings				
spaces	M	L	N	αL	rL
RGB	**77.65**	70.74	71.62	70.12	71.85
L*a*b*	77.76	80.03	77.89	**80.13**	78.02
VSH	77.66	79.23	77.75	**79.43**	77.65
LSH	77.73	80.12	77.74	**80.37**	77.54

Table 3.6: Classification rates in % for the textures of Outex13, using vector erosion based covariance.

realised using a kNN classifier with $k = 1$ [162].

Table 3.6 contains the accuracy rates that have been obtained. In general, the relative performances can be considered similar to those obtained during noise reduction. More precisely, marginal processing continues to outperform its alternatives in RGB, once more followed by norm based ordering. Since all channels are equally important in this colour space, both prioritisation attempts with L and αL fail to surpass it. However, the situation is radically modified with the other three colour spaces, where all approaches increase their performances, a result showing the effect of the colour space choice on their behaviour. Specifically, lexicographical orderings take the lead, though with a small margin, while they provide the overall best results in LSH.

Since brightness alone is considered sufficient for the recognition of most image variants, its prioritisation by means of total orderings has a positive impact on the end results. Similarly to the previous application, phenomenal spaces once more outperform, though only marginally the almost perceptually uniform L*a*b* space. A result which can be deemed natural, since perceptual uniformity is not expected to aid this process in any way. Furthermore, one can also observe the slight superiority of LSH over cylindrical VHS, which could be justified by the multiple structural inconveniencies of the latter.

3.5 Conclusion

This chapter has introduced the theoretical concepts behind the multivariate morphological framework, as well as the various approaches that have been proposed with this purpose. Given the complete lattice theory based structure of morphology, it becomes easy to observe that a vector ordering is sufficient for the definition of valid multivariate morphological operators. That is why besides numerous ordering possibilities, the majority of which are colour specific, have been presented and experimentally tested. Consequently, having studied the overlying application, we now can make a colour space choice, as well as choose a vector/colour ordering family to work with in the sequel of this thesis.

Choosing a colour ordering: The different ordering methodologies that have been elaborated both theoretically and practically, have shown that each is equipped with a distinct set of properties, that render them more appropriate for certain tasks than others. For instance distance based R-orderings provide uniform processing of image channels while lexicographical C-orderings are more suitable for situations with inherent channel prioritisation. Thus, no single approach may be considered as ideal for colour images, since application needs can vary highly. It would not be wise for instance to process in the same way highly saturated or very dark images. Hence, unless a very specific application area with well defined requirements is considered, since the users' needs are not a priori known, *flexibility* on the part of the ordering procedure is a necessity.

Moreover, pre-orderings are not to be considered, since they do not provide the fundamental *anti-symmetry* property. The fact that they can lead to multiple extrema undermines seriously the theoretical soundness of the resulting operators. Additionally, even though certain applications may not require it (e. g. noise reduction), colour preservation is another crucial issue to be taken under consideration in this debate, hence orienting our choice towards total orderings. In particular, among the presented options, we choose to develop our colour morphology framework based upon lexicographical ordering, which besides the aforementioned theoretical advantages, also provides a means of effective customisation through the configuration of the lexicographical cascade's order.

Choosing a colour space: Through the survey provided in this chapter it was observed that several colour spaces have been employed for the definition of colour morphological operators (Table 3.1). Even though the general suitability of a colour space for image processing has been extensively studied by Serra, Hanbury and Angulo [9, 95, 243, 244], we have refrained so far from adopting their view, detailed in Sec. 2.7.1, endorsing non perceptually uniform phenomenal spaces, prior to the study of this particular overlying application.

Our study has shown us that due mainly to their intuitiveness, as far as the human observer is concerned, based on the number of cases mentioned, the use of polar colour spaces appears to be widespread and commonly regarded as "optimal" for this context. Perceptual uniformity on the other hand, as provided by $L^*C^*H^*$, as well as the non-linearity compensation of the lightness component of CIE spaces, are deemed as unnecessary for image processing properties [244], whereas device independence is hardly of interest in generic environments, where more than often no information on the acquisition device's properties are available. Besides, as shown by Eq. (3.3.9) they also offer a control mechanism over the introduction of false colours by limiting the participation of the hue component in ordering. Moreover, as already stated, the notion of flexibility from the part of the ordering under consideration is of importance, hence making a suitable colour choice the one that is able to accommodate this notion. For these reasons, we decide to follow the mainstream trend, and concentrate on phenomenal spaces, and in particular, considering the work of Serra on this topic [243], we opt for the LSH colour space.

Consequently, the next chapter will elaborate on various possibilities concerning the improvement of the existing, as well the development of original and effective colour orderings, based on the lexicographical ordering in the LSH colour space.

Chapter 4

Lexicographical ordering in LSH

4.1 Introduction

Based on the study of colour spaces and multivariate MM, carried out in the previous two chapters, it has been decided to employ in this chapter the lexicographical ordering in combination with the LSH colour space, with the purpose of defining colour morphological operators.

In short, our decision of using a phenomenal colour space has been based on the following arguments: their intuitiveness, as far as human colour perception is concerned, their optional capacity to avoid false colours by ignoring the hue component, as well as their widespread use in colour morphology in the last few years; while the specific choice of LSH, is backed by its corrected saturation formulation, along with improved support for basic colour operations (e. g. averaging, etc). The choice of lexicographical ordering on the other hand, as a means to form a complete lattice in the colour space under consideration, has been made based on its positive theoretical properties, totality and anti-symmetry, in combination with its inherent capacity for flexible ordering configurations.

Consequently, this ordering and colour space may be immediately put to practise, leading to valid colour morphological operators. Albeit this combination of colour space and ordering lead to operators of relative practical interest, a couple of issues requires special attention. These issues are first the processing of the hue component, which demands a method taking into account its periodicity. Moreover, the level of priority attributed to the first vector dimension during lexicographical comparison, hinders highly the effective exploitation of the subsequent image channels. Of course, various solutions for both of these problems have already been proposed, and they have been presented briefly in the survey of the previous chapter.

In this chapter, we will review in larger detail the existing solutions to the aforementioned two problematic aspects of lexicographical ordering in phenomenal colour spaces, as well as their practical limits. In particular, we first elaborate on the issue of hue processing and introduce an approach that makes use of multiple reference hue values, hence leading to more intuitive results (Sec. 4.2). Additionally, new ways of modifying lexicographical ordering for increased symmetry among its components will be presented, including an approach leading to pseudo-morphological operators, all of which are tested through practical applications (Sec. 4.3). Finally, the notions of this chapter are applied to the extension of the hit-or-miss transform, which is defined for colour images (Sec. 4.4).

4.2 Morphological processing of hue

This section focuses on the main inconvenience of phenomenal spaces regarding image processing: the angular nature of hue. Since hue values reside in $[0, 2\pi]$, one can always order them using the scalar order, but there is no a priori reason for setting red ($h = 0$) as "less" than green ($h = 2\pi/3$). The fixed origin and discontinuity at red, besides being an important flexibility constraint, also results in undesirable behaviour in this area of the hue circle. To illustrate this last point, Fig. 4.1 shows the dilation based on the scalar hue order, of a colour image containing all the possible hues. The abrupt discontinuity at red is what causes the edge in the middle of the same figure.

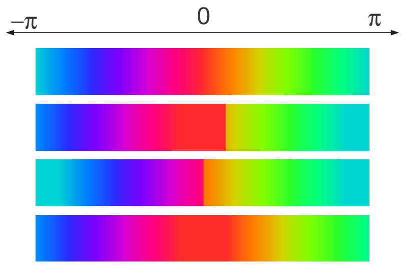


Figure 4.1: From top to bottom, a colour image containing fully saturated hues, its dilation with a 51 pixel wide horizontal line shaped SE based on the scalar order, based on the hue ordering of Eq. (4.2.1) and based on the ordering of Eq. (4.2.3), both with $h_0 = 0.0$.

4.2.1 Mono-reference based hue ordering

Indeed, this situation has led some authors to even ignore this component as in Ref. [273], whereas the first attempt aiming to counter it was made by Peters [207], who employed a circular distance $D(h_0, h)$ measured between the hue of the SE h_0 and the processed value h, with the aim of ordering them:

$$\forall\, h, h' \in [-\pi, \pi[, \;\; h \leq h' \Leftrightarrow D(h_0, h) \leq D(h_0, h') \tag{4.2.1}$$

where $D(\cdot, \cdot)$ is defined as:

$$\forall\, h, h' \in [-\pi, \pi[, \;\; D(h, h') = \begin{cases} h - h' + 2\pi & \text{if} \quad -2\pi \leq h - h' < -\pi, \\ h - h' & \text{if} \quad -\pi \leq h - h' < \pi, \\ h - h' - 2\pi & \text{if} \quad \pi \leq h - h' < 2\pi. \end{cases} \tag{4.2.2}$$

Essentially, since hue values lack an inherent ordering relationship, any arbitrarily developed fixed order (e. g. *green* > *yellow* > *blue*, etc) is of limited practical use. The flexibility of a distance based approach, due to the use of an arbitrary hue origin, renders it adaptive to varying application needs, and that is why this type of hue orderings have met with acceptance. A variant has been later on proposed by Hanbury and Serra [102, 103]. Specifically, according to them, Eq. (4.2.1) results in counter-intuitive operators, since for instance erosions would tend to enlarge objects of the reference h_0 (Fig. 4.1). Hence, they proposed using the ordering:

$$\forall\, h, h' \in [0, 2\pi], \;\; h \leq h' \Leftrightarrow h' \div h_0 \leq h \div h_0 \tag{4.2.3}$$

where, contrary to Eq. (4.2.1), hues closer to h_0 are considered greater and the angular distance $h \div h_0$ of h from h_0, is modified as:

$$\forall\, h, h_0 \in [0, 2\pi], \quad h \div h_0 = \begin{cases} |h - h_0| & \text{if} \quad |h - h_0| < \pi \\ 2\pi - |h - h_0| & \text{if} \quad |h - h_0| \geq \pi \end{cases} \tag{4.2.4}$$

For normalised hue values $h \in [0, 1]$, it suffices to replace π with 0.5 in Eq. (4.2.4). Fig. 4.1 presents the result of a hue dilation based on the ordering of Eq. (4.2.3). As vectors are ordered with respect to their distance from $h_0 = 0.0$, the edge appearing in Fig. 4.1 is avoided within the red region. Furthermore, when combined with unsupervised reference hue computation methods [6, 96], the ordering of Eq. (4.2.3) can be effectively used for general purpose morphological hue processing [103]. Let us note however that the term "ordering" is used loosely in the context of distance based reference hue computations, since they obviously do not verify the anti-symmetry constraint, as two distinct hue values can lie at equal distances to the reference hue; hence they constitute pre-orderings to be exact. One possible approach for satisfying the desired anti-symmetry property, is to use an arbitrary ordering direction within the hue circle with the purpose of resolving such conflicts [104]. For instance if $h \div h_0 = h' \div h_0$ and $h \neq h'$, then the hue located first in the chosen ordering direction is considered smaller.

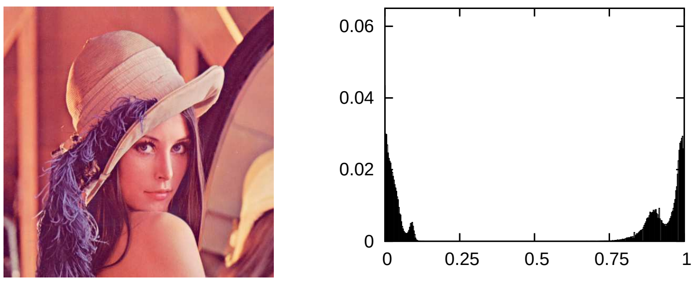


Figure 4.2: The image Lenna (left) and its normalised hue histogram (right).

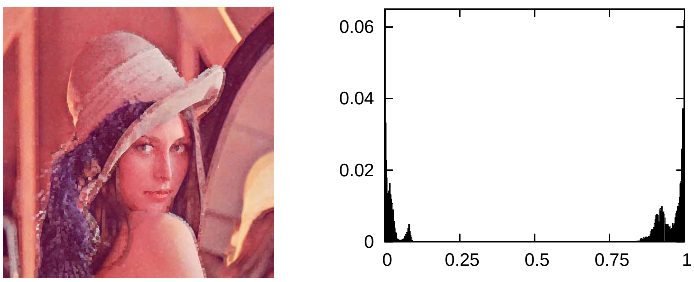


Figure 4.3: The image Lenna (left), after the hue channel has been dilated with a square shaped SE of size 7×7 pixels using $h_0 = 0.0$ and ordering of Eq. (4.2.3), and the corresponding normalised hue histogram (right). Observe the concentration around h_0.

4.2.2 Multi-reference based hue ordering

Using a variable hue origin is certainly a practical way of introducing a complete lattice structure on the hue circle. However, as it will be subsequently detailed, this approach has certain limitations, the effects of which depend on the image at hand. Here we propose a more flexible new hue ordering based on distances from multiple origins. For the sake of simplicity, in the sequel all hue values are normalised to $[0, 1]$.

Let us consider the case of the Lenna image (Fig. 4.2 (left)). As it can be seen from its hue histogram (Fig. 4.2 (right)), there is a high concentration around the origin of the hue circle. And since reference hues are chosen usually as the average hue or the most frequent [6, 96], it is considered pertinent to choose $h_0 = 0.0$. By accepting a certain wavelength within the red region of the spectrum as the reference hue, the nuances of red closer to this value increase their presence within the image by means of a dilation (Fig. 4.3). In other words "red $>$ orange and red $>$ magenta". But how about blue, and green or cyan? Their relationship to the reference hue is *implicitly* assumed to be that of their position on the hue circle. Specifically, red is equidistant to blue and green, and furthest from cyan, based on the principle of colour complementarity.

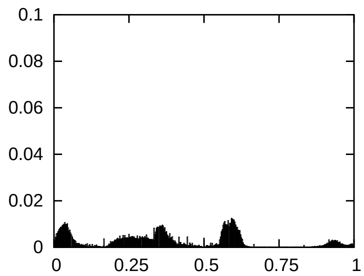


Figure 4.4: The image Cat (by Gamze Aktan) (left) and its normalised hue histogram.

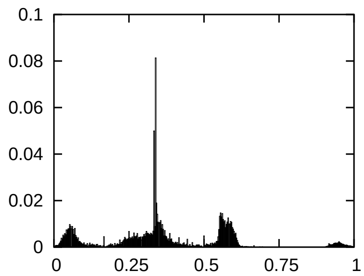


Figure 4.5: The image Cat (left), after the hue channel has been dilated with a square shaped SE of size 7×7 pixels using $h_0 = 0.34$ and ordering of Eq. (4.2.3), and its corresponding hue histogram (right); note that hue values are normalised to $[0, 1]$.

This approach however assumes that the input image has at most a single dominant hue, which in the case of Lenna is red. Hence, in order to make the connection with binary morphology, the reference hue h_0 represents the "foreground" and its complement on the hue circle the "background". However, in the general case the presence of more than one not dominant hues is rather more frequent. Observe the image Cat (Fig. 4.4 (left))

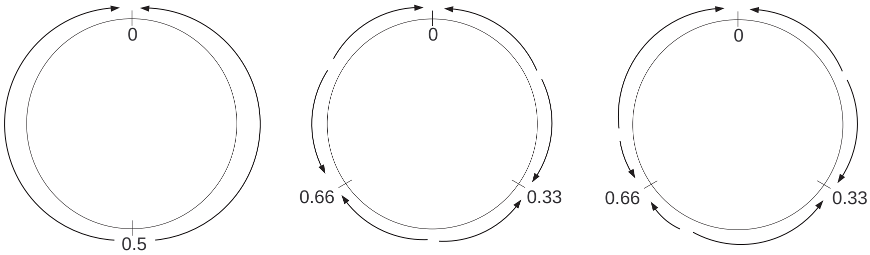


Figure 4.6: From left to right, the hue circle with arrows showing the direction of increasing hues, in the case of a single reference hue ($h_0 = 0.0$), with multiple reference hues ($h_0 = 0.0$, $h_1 = 0.33$, $h_2 = 0.66$), and with multiple reference hues weighted according to arbitrary significances.

for instance, where primarily three colours appear to be frequent: red/orange, green and blue, corresponding to the three large peaks in its hue histogram (Fig. 4.4 (right)). The reference hue that was obtained ($h_0 = 0.34$) through averaging, shows that green-yellow is the dominant wavelength, as it occupies a relatively important portion of the total image area. Consequently, during the dilation of the same image (Fig. 4.5), hues are ordered only based on their distance from green, even though orange and blue correspond to almost equally important regions/objects within the image.

To counter this situation, one could eventually employ a region specific hue reference, for instance by computing h_0 independently for each hue region. And during ordering, the hues of a given region could be ordered according to their distances from the local reference hue. Nevertheless, this approach is hard to use in practice as it requires a pertinent segmentation step, that will provide the necessary region borders. Moreover, the image content may be of such nature that regions may not even be distinguishable, and furthermore, how to order hues that belong to different regions? Alternatively, one can also adopt a pixel specific approach, where the reference hue is constantly updated according to the pixels under the SE. This use of a variable reference hue however, undermines several theoretical properties of the resulting operators, since "distant" pixels become no longer comparable.

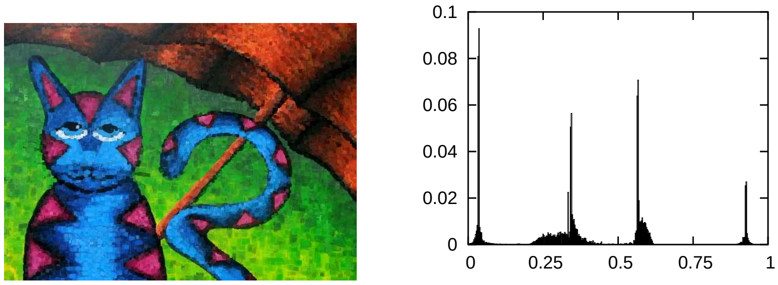


Figure 4.7: The image Cat (left), after the hue channel has been dilated with a square shaped SE of size 7×7 pixels using ordering of Eq. (4.2.5), references $h_0 = 0.04, h_1 = 0.34, h_2 = 0.56, h_3 = 0.93$, and its corresponding normalised hue histogram (right).

The solution we propose here, is to use a hue ordering based on multiple references. Specifically, let us assume that we have obtained the k representative hue values

$R = \{h_i\}_{1 \leq i \leq k}$. In the case of the image Cat (Fig. 4.4 (left)), the set R would contain roughly the hues red/orange, green/yellow, blue and magenta. Given two hue values, we can now order them based on their distances from each of the k references, and choose as greater the one closer to a representative hue value. More precisely:

$$R = \{h_i\}_{1 \leq i \leq k}, \quad \forall\, h, h' \in [0,1], \quad h \leq_H h' \Leftrightarrow \min_i \left\{h' \div h_i\right\} \leq \min_i \left\{h \div h_i\right\} \tag{4.2.5}$$

Consequently, if the set R contains a single reference hue, such as in the case of the image Lenna, then Eq. (4.2.5) becomes obviously equivalent to Eq. (4.2.3) (Fig. 4.6 (left)). However, if on the contrary R contains multiple references, then during ordering we favour those hues that are closer to one of the representative references (Fig. 4.6 (middle)), hence dividing the hue circle into regions with each having its own extremum. In the case of the Cat image, given two nuances of green, the greater would be the one which is closer to the reference hue situated in the green region of the hue circle, whereas if we compare a nuance of blue against one of green, the one closer to its respective reference would be considered greater, hence favouring "representative" hues, in terms of their distance from their respective reference hues. This can be easily observed by examining the hue histogram of the image Cat, obtained after having applied a dilation based on this principle (Fig. 4.7 (right)). The concentration of hue values around the four major hues is obvious.

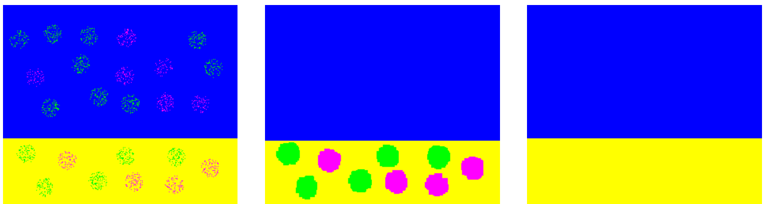


Figure 4.8: From left to right, the original image, its hue dilation using the ordering of Eq. (4.2.3) and $h_0 = 0.66$, and the proposed ordering of Eq. (4.2.5) with $h_0 = 0.66$, $h_1 = 0.33$, both with a square shaped SE of size 5×5 pixels.

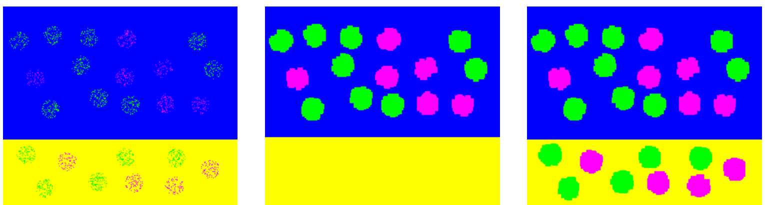


Figure 4.9: From left to right, the original image, its hue erosion using the ordering of Eq. (4.2.3) and $h_0 = 0.66$, and the proposed ordering of Eq. (4.2.5) with $h_0 = 0.66$, $h_1 = 0.33$, both with a square shaped SE of size 5×5 pixels.

Since it is relatively hard to detect the visual differences among the previous images, consider the artificial example shown in Fig. 4.8 (left). It consists of two uniform blue and yellow regions, containing noise spots of green and magenta. The "most dominant" hue is evidently blue, hence, a dilation using this single reference leads to the removal of spots only from the upper area (Fig. 4.8 (middle)), since both green and magenta are closer to blue than yellow. On the other hand, both regions become noise free using the proposed approach in combination with the two dominant hues, i. e. blue and yellow (Fig. 4.8 (right)).

Similarly, in the case of an erosion, non-representative hues, in other words hues that are distant from the reference points on the hue circle are favoured. Consequently, the approach based on a single reference erodes the spots in the yellow region, since both green and magenta are closer to yellow than blue (Fig. 4.9 (left)). A more consistent behaviour is achieved with the proposed ordering (Fig. 4.9 (right)).

Then again, likewise to Eq. (4.2.3), $\leq_H$ is too a pre-ordering, and unfortunately we can no longer use the arbitrary hue circle direction to resolve equivalences among distinct hues [104], as described in Sec. 4.2.1, since now we have multiple reference hues. That is why an alternative approach is required in order to satisfy the anti-symmetry property. One way of achieving this goal is to employ the order of importance among references, so that the hue closer to a "more important" reference hue is considered greater, and if that does not resolve the ambiguity, comparing the scalar hue values certainly does. Specifically:

$$h <_H h' \Leftrightarrow \begin{cases} \min_i \{h' \div h_i\} < \min_i \{h \div h_i\}, \text{ or} \\ \min_i \{h' \div h_i\} = \min_i \{h \div h_i\} \text{ and } f(\mathrm{ref}(h)) < f(\mathrm{ref}(h')), \text{ or} \\ \min_i \{h' \div h_i\} = \min_i \{h \div h_i\} \text{ and } f(\mathrm{ref}(h)) = f(\mathrm{ref}(h')) \text{ and } h < h' \end{cases} \tag{4.2.6}$$

where $\mathrm{ref}(h) = \mathrm{argmin}_{h_i \in R} \{h \div h_i\}$ is the reference hue closest to a hue value h and $f(\cdot) : [0, 1] \to \mathbb{R}$ is the function associating to a reference hue its relative "importance" (detailed further later in this section).

Of course, a necessary step for the implementation of the hue ordering in Eq. (4.2.5), consists in obtaining the reference hue(s). Moreover, unless an expert is available, it needs to be carried out automatically. Finding the "important" histogram maxima, can be considered as a histogram clustering or segmentation problem. As such, several options can be used with this purpose, for example any of the known unsupervised clustering techniques (e. g. meanshift), or even a morphological watershed transform. Through empirical means, the following method was observed to provide robust results:

1. Calculate the normalised hue histogram, where bin h_i is denoted by $P(h_i)$

2. Threshold the histogram using its average

3. Find the connected histogram sections with non-zero bins

4. For each section between bins h_i and h_j, calculate its maximum, as reference hue:

$$h_{max} = \arg \max_{h_k \in [h_i, h_j]} P(h_k) \tag{4.2.7}$$

However, using this reference computation scheme, one can eventually obtain several histogram sections, each with its own maximum, hence resulting in an excessively subdivided hue circle. To counter this eventual undesirable situation, one can of course first filter the histogram, using for instance an opening with a horizontal SE, but a more efficient way consists in weighting the distance from each reference hue, with the "importance" of its originating histogram section. To illustrate this idea, consider the histogram of the Cat image (Fig. 4.4), which with the proposed reference hue computation method, results in four sections, representing roughly the hues red/orange, green, blue and magenta, with the last being of relatively lesser significance. In order to quantify the "importance" of the i^{th} section ($1 \leq i \leq m$) one can use either the length (l_i) of the sector or preferably its discrete sum (w_i). Thus we can modify the ordering of Eq. (4.2.5) to accommodate the significance n_i of each reference hue h_i:

$$R = \{(h_i, n_i)\}_{1 \leq i \leq k}, \forall\, h, h' \in [0,1],$$
$$h \leq h' \Leftrightarrow \min_i \left\{(h' \div h_i) \times 1/n_i\right\} \leq \min_i \left\{(h \div h_i) \times 1/n_i\right\} \quad (4.2.8)$$

where n_i is defined as:

$$n_i = \frac{w_i}{\sum_{k=1}^{m} w_k} \quad (4.2.9)$$

By means of the factor n_i, we ensure that hues that are moderately close to a significant reference hue, are considered larger than those much closer to a reference hue of lesser importance, hence providing increased robustness against noisy histograms (Fig. 4.6 (right)). Note that n_i may also be used as the reference importance function in Eq. (4.2.6). If we return to our example, this approach leads to the decrease of magenta hues after dilation (Fig. 4.10), since they are much less frequent than the other three major hues.

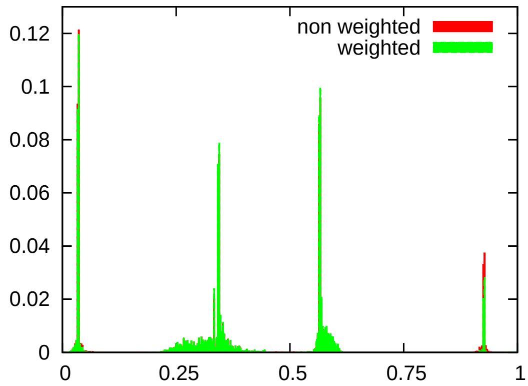


Figure 4.10: The hue histograms of image Cat, after applying a dilation using orderings of Eqs. (4.2.5) and (4.2.8) along with references $h_0 = 0.04$, $h_1 = 0.34$, $h_2 = 0.56$, $h_3 = 0.93$, both with a square shaped SE of size 11×11 pixels. This figure is best viewed in colour.

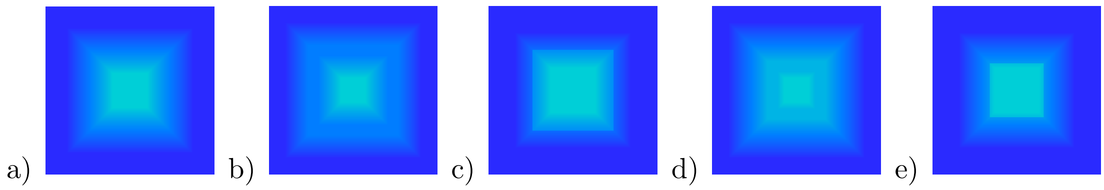


Figure 4.11: From left to right, the original image, its hue erosion (b,d) and dilation (c,e) using orderings of Eqs. (4.2.5) and (4.2.8), both with references $h_0 = 0.66$, $h_1 = 0.5$ and their respective weights $n_1 = 0.8$ and $n_2 = 0.2$ along with a square shaped SE of size 25×25 pixels.

A further example demonstrating the effect of the proposed orderings, as well as the difference between Eqs. (4.2.5) and (4.2.8) is shown in Fig. 4.11. The original image

(Fig. 4.11 (left)), contains a uniform hue gradient between a central cyan square and a larger surrounding blue area. Using the aforementioned reference computation scheme based on the hue histogram, we obtain as references the two values representing cyan and blue. For the sake of clarity, Fig. 4.12 shows additionally the uni-dimensional version of the hue gradient and the effect of the erosion and dilation operators, both with and without the use of weights. A relatively large SE has been used, in order to amplify the effect of each operator.

The dilation operator, applied based on Eq. (4.2.5) (Fig. 4.12 (left)), results in the discretisation of the gradient, by increasing each representative hue's influence range, hence creating the steep edge in the middle, showing the limit of each reference hue's attraction range (Fig. 4.6 middle). In other words, "representative" hues are favoured while the intermediate values are eliminated. Erosion on the other hand, favours the intermediate hue values between the two references, thus leading to the creation of a "middle" flat zone, and forcing the two gradient extremes edges to shrink.

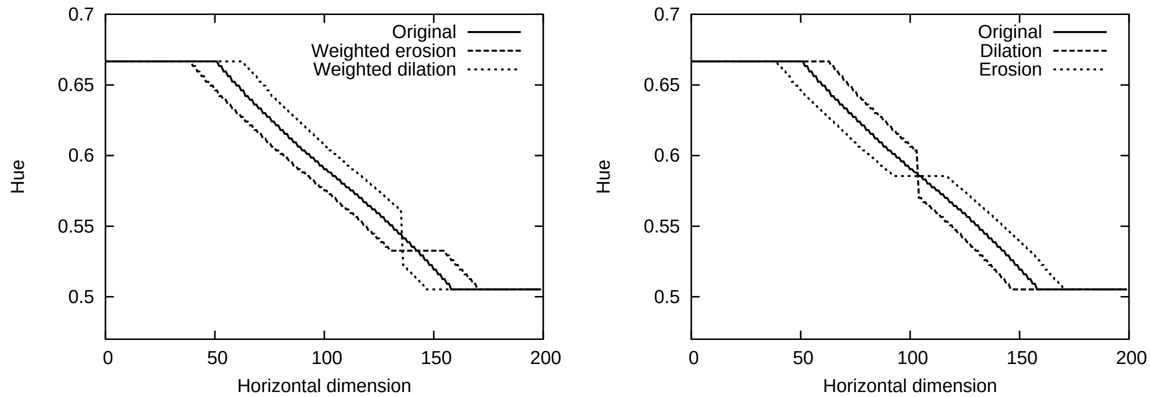


Figure 4.12: The uni-dimensional gradient of Fig. 4.11, along with its dilated and eroded versions with (right) and without (left) using weights.

By taking into account the much smaller size of the central cyan area with respect to the outer blue, we obtain the results depicted in Fig. 4.12 (right). While dilation and erosion behave identically as in the unweighted case, the influence range of each reference hue is radically modified in favour of the obviously more dominant blue. Consequently, given a noisy hue channel which could lead to the eventual unsupervised detection of unimportant hue references, the weights incorporated within the ordering of Eq. (4.2.8) can decrease substantially their influence on the end result.

Prior to any practical use however, one last point that remains is the disturbance caused within the histogram by hue values corresponding to achromatic pixels. This problem may be resolved by using a function making it possible to distinguish pixels of chromatic value from those that are achromatic. Of course this is no deterministic procedure, and a continuous model is necessary. We choose to employ a trivial linear model, where the value of saturation provides a direct indication on the importance of hue. Specifically, given a coefficient $c(x, y) \in [0, 1]$ (i.e. the saturation at x, y) indicating the colourfulness level of a colour pixel at position x, y, the histogram can be constructed by incrementing the corresponding bin by c instead of 1. In other words, the total value in bin h_i of the hue histogram, is calculated as:

$$P(h_i) = \sum_{x,y} c(x,y)\delta_{h_i h_{x,y}} \tag{4.2.10}$$

where the sum is over all the pixels at x, y, $h_{x,y}$ is the hue value at this position and δ_{ij} denotes the Kronecker delta function. In the sequel of this thesis, wherever mentioned, the hue channel will be being processed according to Eq. (4.2.8) and the conflict resolution scheme of Eq. (4.2.6) using the reference hues computed according to the aforementioned threshold based approach, through histograms constructed by means of Eq. (4.2.10), unless specified otherwise.

Figure 4.13: The test images, from left to right, Lenna, Macaws and Mandrill.

Images	Equalities (%)	Red (%)	Green (%)	Blue (%)
Lenna	0.12	93.02	6.3	0.56
Macaws	8.89	83.3	5.71	2.1
Mandrill	0.54	95.28	3.51	0.67

Table 4.1: The percentages of comparisons that have led to equalities, and those that have been determined at each channel, during the dilation of the images of Fig. 4.2.2, with a lexicographical ordering ($R \rightarrow G \rightarrow B$) using a square shaped SE of size 5×5 pixels.

4.3 Lexicographical ordering

Lexicographical ordering, defined in Eq. (3.2.6), as explained in Sec. 3.3.4, possesses multiple properties than render it a more adequate choice than its competitors, as a way of ordering vectors in the context of multivariate MM. Specifically, the first property is its totality. Since all vectors are lexicographically comparable, according to Ref. [258] it preserves the original vectors of the input image, hence preventing the appearance of false colours. Moreover, as it satisfies the anti-symmetry constraint, a property lacking from most reduced orderings, it makes it possible to compute unique vector extrema, thus effectively avoiding ambiguities during vector ordering. Additionally, it represents an excellent choice in combination with spaces possessing an either inherent or artificially induced assymetric distribution of information; which is asserted by the number of authors that have considered it (Table 3.1), due to its capacity to attribute a certain amount of priority to the first vector or colour dimension. That is why it has been popular with phenomenal spaces, that separate effectively the different notions of higher level colour perception. Additionally, through the control of the channels' order during comparison (e. g. luminance first, saturation second, etc), it provides a certain degree of customisation flexibility, a notion of crucial importance as far as general purpose colour morphological operators are concerned.

However, notwithstanding its desirable properties, lexicographical ordering also suffers from a serious drawback. More precisely, the outcome of the vast majority of lexicographical comparisons, is decided based only on the first few vector components that are compared, while the contribution of its remaining dimensions is typically negligible. This property is illustrated in Table 4.1, where the percentages of comparisons determined by the three channels of the three RGB colour images of Fig. 4.2.2, during a vector dilation based on a lexicographical ordering are shown. The channel occupying the first position of the lexicographical cascade (i. e. red) is obviously responsible for the vast majority of comparisons outcomes.

Of course this might be a desired behaviour in cases where the fist image channel contains the majority of the total variational information, for instance after applying a PCA transform. Nevertheless, most often it leads to an insufficient exploitation of the image channels and inter-channel relations. This effect is most aggravated in the case of hyperspectral images, where despite the availability of hundreds of channels, only at most the first few participate in the overall process. That is why, variations of Eq. (3.2.6) have been proposed, with the purpose of better tuning the priority as well as degree of influence of each vector component on the comparison outcome, by means of an user defined argument α.

The first attempt aiming to decrease the priority attributed to the first vector component during lexicographical comparison was made by Ortiz et al. [198], that proposed the *α-lexicographical ordering*:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \;\; \mathbf{v} < \mathbf{v}' \Leftrightarrow \begin{cases} v_1 + \alpha < v_1', \;\text{ or} \\ v_1 + \alpha \geq v_1' \;\text{ and }\; [v_2, \ldots, v_n]^T <_L [v_2', \ldots, v_n']^T \end{cases} \tag{4.3.1}$$

where $\alpha \in \mathbb{R}^+$. The α argument is thus used to the end of increasing the occurrence of equivalences within the first vector dimension, since a scalar value v_1 becomes "equal" to all values contained in the interval $[v_1 - \alpha, v_1 + \alpha]$, hence allowing comparisons to reach more frequently the second dimension. Nevertheless, Eq. (4.3.1) is not transitive, and consequently does not represent an ordering from an algebraic point of view.

A theoretically sounder approach was proposed by Angulo and Serra [11] and was named *α-modulus lexicographical ordering*:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \;\; \mathbf{v} < \mathbf{v}' \Leftrightarrow [\lceil v_1/\alpha \rceil, v_2, \ldots, v_n]^T <_L \left[\lceil v_1'/\alpha \rceil, v_2', \ldots, v_n'\right]^T \tag{4.3.2}$$

which aims to create equivalence groups within the first dimension. It relies on a quantisation through division by a constant α followed by a rounding off, which reduces the dynamic margin of the first dimension, thus allowing a greater number of comparisons to reach the second. On the contrary of the aforementioned approach, it is reflexive and transitive, hence a pre-ordering.

The effect of the operation realised on the first dimension, becomes more clear by studying the *space filling curves* (SFC), that travel through points of multi-dimensional space. Since the search for a total vector ordering can be formulated as a search for an injective function mapping all the points of a multi-dimensional space onto a uni-dimensional space, SFC satisfy this requirement and make it possible to model different solutions, where vectors are ordered according to the position of their coordinates on the SFC. Since lexicographical ordering corresponds to a bijection [40], its SFC will pass once on all points of a bi-dimensional

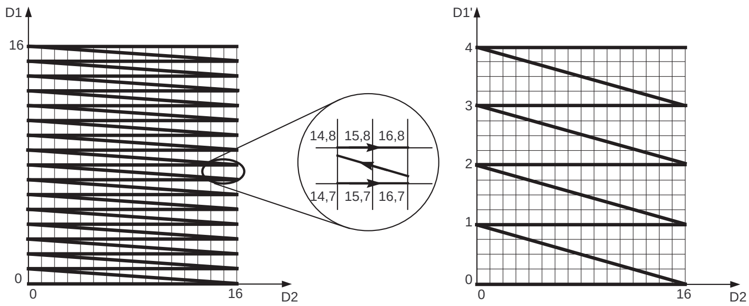


Figure 4.14: Space filling curves in a bi-dimensional space for a lexicographical ordering (left), where the arrows denote the direction of increasing vector coordinates, and for the α-modulus lexicographical ordering with $\alpha = 4$ (right).

discrete space $D1 \times D2 = [0, 16]^2$ as illustrated in Fig. 4.14 (left), where the high priority attributed to the first dimension ($D1$) is represented by the high frequency of horizontal curves. Fig. 4.14 (right) shows the quantised form $D1'$ of dimension $D1$ with $\alpha = 4$, which leads to the creation of equivalence groups. For instance the points $\{5, 6, 7, 8\}$ of $D1$ now belong to the same group 2 of $D1'$, and thus given two coordinates $(5, 1)$ and $(7, 0)$ in $D1 \times D2$, the first components are considered equal in $D1' \times D2$ and the outcome of the comparison is determined by the second dimension, in other words $(5, 1) > (7, 0)$.

Nevertheless, special attention is required with the use of this approach, since the resulting equivalence groups obviously eliminate the anti-symmetry property of lexicographical ordering. One way of countering this problem, that renders this approach as well as any other similar to it anti-symmetric and thus an ordering, is to continue the lexicographical cascade with the un-quantised vector dimensions. To explain :

$$\forall \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \;\; \mathbf{v} < \mathbf{v}' \Leftrightarrow \begin{cases} [j_1(v_1), \ldots, j_{n-1}(v_{n-1}), v_n]^T <_L \left[j_1(v'_1), \ldots, j_{n-1}(v'_{n-1}), v'_n\right]^T, \text{ or} \\ [j_1(v_1), \ldots, j_{n-1}(v_{n-1}), v_n]^T = \left[j_1(v'_1), \ldots, j_{n-1}(v'_{n-1}), v'_n\right]^T \\ \text{and } \mathbf{v} <_L \mathbf{v}' \end{cases} \tag{4.3.3}$$

where $j_i(\cdot),\; i \in [1, n-1]$ denotes some function used to reduce the dynamic margin of the i^{th} dimension. Thus only equal vectors are considered equivalent.

In conclusion, among the solutions that have been so far reported with the purpose of fine-tuning the dimension prioritisation of lexicographical ordering, only Eq. (4.3.2) in combination with Eq. (4.3.3) satisfies the theoretical requirements of an ordering and hence leads to valid morphological operators. The practical use of α-modulus lexicographical ordering however is limited, as it relies on an implicit assumption on the dimension to be quantised. The next sections introduce a generalisation of this approach that makes it possible to employ a wider range of quantisation functions, as well as two other ways for decreasing the priority attributed to the first image channel by lexicographical ordering.

4.3.1 Quantisation based α-lexicographical ordering

As far as its theoretical properties are concerned, α-modulus lexicographical ordering appears to be the most pertinent among the two variations, as it provides an effective means of shifting

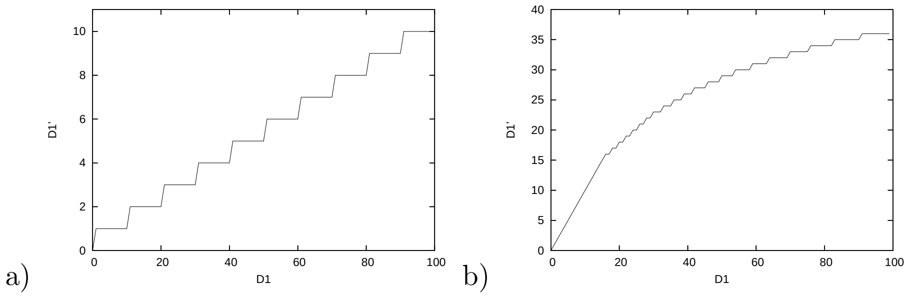


Figure 4.15: Example of quantisation applied on a dimension D1 using $\alpha = 10$ and Eq. (4.3.2) (a), and using the proposed approach with the same α and an exponential distribution (b).

priority away from the first vector dimension, while preserving the desirable characteristics of lexicographical ordering that have made it popular within the colour morphological context.

However, its practical use relies on an important implicit assumption. More precisely, as shown in Fig. 4.15a, the dimension under consideration is processed uniformly, based on the assumption that all of its subsets are equally "unimportant" with respect to the second dimension in the lexicographical cascade. In practice, often complicated relations are present among the image channels, and the need to shift priority to the next dimensions arises only partially. Consider for instance the case of colour morphology in a phenomenal colour space of type LSH, where the importance of saturation is minimised for extreme levels of luminance.

Algorithm 1 The algorithm for computing a quantised discrete dimension based on an arbitrary priority distribution.

Input: $I = [a, \dots, b] \subseteq \mathbb{Z}$, an array containing
the discrete dimension to have its dynamic margin reduced
$\alpha \in \mathbb{N}^*$, a parameter setting the maximum allowed size
of equivalence groups within I
$f : I \to [0, 1]$, a function modelling the desired priority distribution within I
Output: $J \subseteq \mathbb{N}$, the array containing the new quantised dimension
$tmp \leftarrow 0$
for $i \leftarrow a$ to b **do**
  $k \leftarrow \lceil \alpha \times f(I[i - a]) \rceil$
  for $j \leftarrow i$ to $i + k$ **do**
    $J[j - a] \leftarrow tmp$
  end for
  $tmp \leftarrow tmp + 1$
  $i \leftarrow i + k$
end for

In the light of these remarks, we propose the generalisation described in Algorithm 1, as a replacement for the division followed by rounding off, of Eq. (4.3.2). Based on a simple principle, this algorithm realises a quantisation of a given discrete "pixel range" interval, in practice often $I = [0, 255]$, by associating to each value, a group of equivalence, the size of which is computed using a user defined function f, and is limited by the value of $\alpha \in \mathbb{N}^*$. Hence, it is possible to reformulate Eq. (4.3.2) as:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \ \ \mathbf{v} < \mathbf{v}' \Leftrightarrow [w_1, v_2, \dots, v_n]^T <_L [w_1', v_2', \dots, v_n']^T \tag{4.3.4}$$

where w_1 and w_1' represent respectively the equivalence group of v_1 and v_1', obtained through Algorithm 1. Of course one is by no means limited to applying this procedure only to the first dimension. In fact, complicated inter-channel relations often require the repeated use of such an approach in more than one dimensions in order to be effectively modelled.

Similar to Eq. (4.3.2), Eq. (4.3.4) also represents a relation lacking the anti-symmetry property for the same reasons. That is why from this point on all practical applications using Eq. (4.3.4) are to be considered using implicitly its combination with the standard lexicographical ordering, as shown in Eq. (4.3.3), for tie-breaking purposes.

An example of this algorithm is given in Fig. 4.15b, where it is applied on the integer interval $[0, 100]$ using an exponential priority distribution. In other words, values close to zero are of high importance and need to be processed with a fine precision whereas the second dimension may be used with values approaching 100. The increase in the size of equivalence groups can be easily observed as the values approach the upper interval bound. Consequently, in order to obtain the uniform distribution of Fig. 4.15a, corresponding to the quantisation of α-modulus lexicographical ordering, with this approach, it is sufficient to employ the constant function $\forall n \in \mathbb{N}$, $f(n) = 1$ within Algorithm 1. By means of the function f, once can thus model arbitrary priority distributions within image channels. Furthermore, image-specific ordering approaches may be developed using the histogram of the dimension under consideration, as shown in Fig. 4.16, hence leading to adaptive vector orderings.

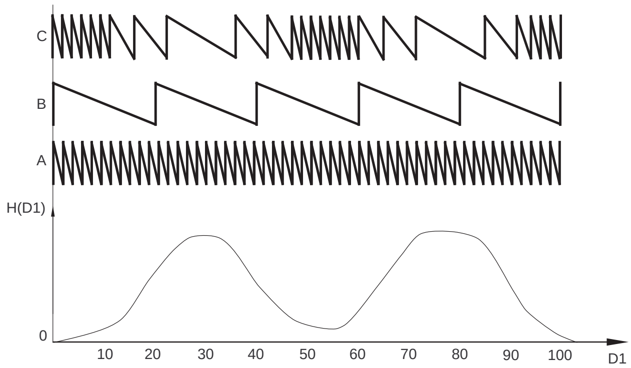


Figure 4.16: The probability distribution of dimension D1 (bottom) for an arbitrary image, general trend of the space filling curve frequency corresponding to its lexicographical ordering (A), α-modulus lexicographical ordering with $\alpha = 20$ (B), and quantisation based α-lexicographical ordering using the dimension histogram as priority distribution function (C).

As a more realistic example, let us return to the aforementioned case of colour morphology in the LSH colour space. Polar colour spaces exhibit an inherent prioritisation of channels, as they model the higher levels of human colour vision, and thus luminance is most often used in the first position of the lexicographical cascade when developing morphological operators [8, 157, 198]. However, non trivial relations are present among the three channels of the bi-conical LSH space, that are totally ignored using Eq. (4.3.2). Specifically:

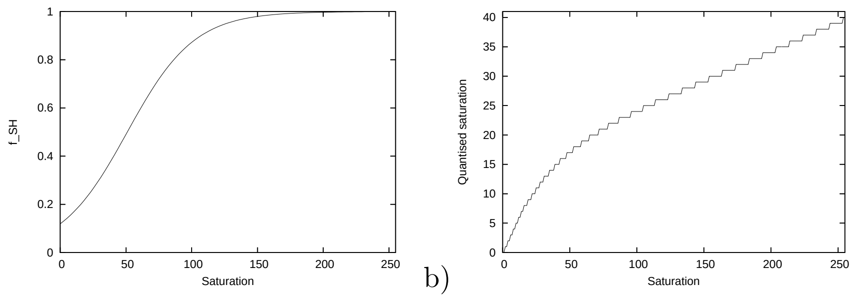
- the importance of saturation is maximised for "medium" levels of luminance,

• whereas hue is of practical importance only for relatively high levels of saturation.

Consequently, one can take into consideration the aforementioned relations through the proposed quantisation scheme, by simply choosing an adequate f function in order to model them. Nevertheless, to the best of our knowledge, there is no definite model for representing these relations in a colour space of this type. Carron [32] for instance has worked on the problem of segmentation along with phenomenal spaces, and has chosen to employ a sigmoidal model for the relation of hue and saturation. Here, as the objective of this thesis is not of colorimetric nature, we adopt his proposal and use it also with the purpose of modelling the relation between luminance and saturation. In particular, the function f_{SH} used for modelling the importance of hue values depending on saturation is as follows:

$$\forall\, s \in [0, 255] \;\; f_{SH}(s) = \frac{1}{1 + exp(-k(s - s_0))} \tag{4.3.5}$$

the form of which is illustrated in Fig. 4.17a and the arguments have been set as $k = 0.04$ and $s_0 = 50$ [32]. Hence, large equivalence groups are formed for relatively saturated colours, facilitating in that case the transition of lexicographical comparisons to hue, as shown in Fig. 4.17b. On the other hand, for low values of saturation, smaller equivalence groups are formed, thus rendering it more difficult to obtain an equality of saturations.



a) b)

Figure 4.17: A model of the "importance" of hue with respect to saturation in the LSH colour space with a sigmoidal distribution (a), and the quantised form of the saturation channel using this relation (b) with $\alpha = 10$ and Algorithm 1.

As to the couple of luminance and saturation, we choose once again a sigmoidal approach, but this time in order to adapt to the bi-conical form of the LSH space, a double sigmoid represents a more adequate choice:

$$\forall\, l \in [0, 255], \;\; f_{LS}(l) = \begin{cases} \frac{1}{1+exp(-m(l-l_l))} & \text{if } l \leq 0.5 \\ \frac{1}{1+exp(m(l-l_u))} & \text{if } l \geq 0.5 \end{cases} \tag{4.3.6}$$

where $m = 0.07$, $l_l = 64$ and $l_u = 192$. The form of f_{LS} as well as the corresponding equivalence groups are given respectively in Figs. 4.18a and 4.18b. Thus, lexicographical comparisons tend to be shifted towards the second dimension, i. e. saturation, for relatively medium values of luminance, which correspond to the middle region of the cone. Even though our model choice for this relation has no colorimetric foundation, we believe that it approximates our intuitive understanding of it sufficiently well to be of practical use. Of course, when a better model is proposed in the future, it can be directly used in combination

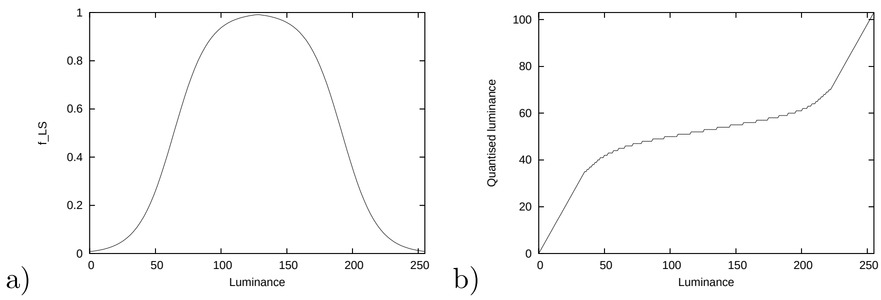


Figure 4.18: A model of the "importance" of saturation with respect to luminance in the LSH colour space with a double sigmoidal distribution (a), and the quantised form of the luminance channel using this relation (b) with $\alpha = 10$ and Algorithm 1.

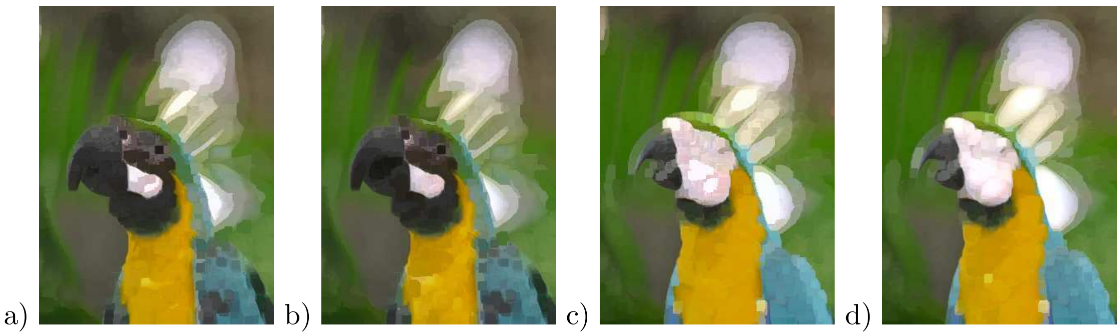


Figure 4.19: The erosion and dilation results of a part of the macaws image (Fig. 4.22a) with a square shaped SE of size 9×9 pixels, using α-modulus lexicographical ordering, Eq. (4.3.2) (a,c) and the proposed quantisation based lexicographical ordering, Eq. (4.3.7) (b,d) both with $\alpha = 50$.

with Algorithm 1, which can accommodate it trivially. Consequently, the resulting ordering, specific to colours of the LSH space becomes:

$$\forall\, \mathbf{v}_1 = (l_1, s_1, h_1), \mathbf{v}_2 = (l_2, s_2, h_2) \in [0, 255]^3, \;\; \mathbf{v}_1 < \mathbf{v}_2 \Leftrightarrow \left[l'_1, s'_1, h_1\right]^T <_L \left[l'_2, s'_2, h_2,\right]^T \tag{4.3.7}$$

where l'_1, l'_2 denote respectively the quantised forms of l_1, l_2 according to Algorithm 1 and priority model f_{LS} of Eq. (4.3.6), while s'_1, s'_2 represent respectively the quantised forms of s_1, s_2 with the model f_{SH} of Eq. (4.3.5). Hence, we manage to render the lexicographical ordering capable of accommodating the inner relations present among the three channels of LSH.

An application example is given in Fig. 4.19, where the results obtained after dilation and erosion with orderings of Eqs. (4.3.2) and (4.3.7) using the same value of α are shown. The differences between the two approaches are fairly visible, in particular at the beak and head region, where more smoothed variations are obtained. As illustrated by this relatively simple example, the proposed priority reduction scheme constitutes a generalisation of the quantisation step in Eq. (4.3.2), that provides a high degree of flexibility, making it possible for the user to accommodate within the resulting ordering relation practically any particular model of priority distribution.

4.3.2 Marker based lexicographical ordering

Creating artificial equivalence groups within the dimension to have its lexicographical priority reduced, is an effective means of shifting priority to the remaining vector dimensions. However, this is an operation realised independently from the images to be processed (except if the priority distribution is image specific, e. g. based on its histogram), and no a priori information is available on where the, eventually abrupt, equivalence group limits correspond within the input image. This undesirable situation is illustrated in Fig. 4.20 where the equivalence groups of a uni-dimensional discrete signal, obtained with Eq. (4.3.2), lead to artificial edges/value variations during comparison.

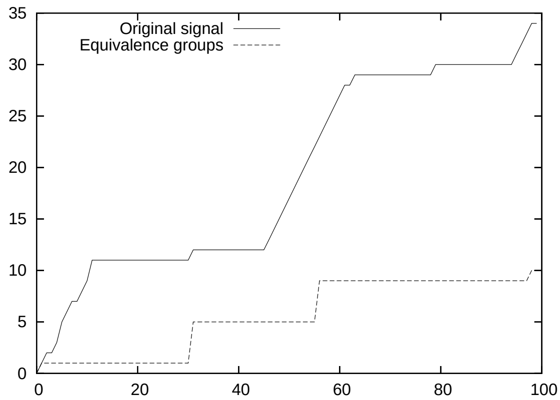


Figure 4.20: Example of a uni-dimensional discrete signal with its corresponding equivalence groups obtained with Eq. (4.3.2) and $\alpha = 11$.

An alternative way of achieving a priority shift that avoids this inconvenience, consists in assuming an image specific approach and forming these equivalence groups based on the content of the image to be processed. Specifically, one can preprocess the input image's channel that is to have its priority reduced, for instance the luminance channel in the example of Sec. 4.3.1, so that it is "flattened", or in other words heavily smoothed. If we call this image a "marker" m, then the ordering of vectors of an image g can be realised as follows:

$$\forall x, y, s, t \in \mathbb{Z}, \;\; g(x,y) < g(s,t) \Leftrightarrow \\ [m(x,y), g_2(x,y), \ldots, g_n(x,y)]^T <_L [m(s,t), g_2(s,t), \ldots, g_n(s,t)]^T \quad (4.3.8)$$

In other words, the formation of equivalence groups is now controlled only by the marker image. If a couple of pixels have the same value in the marker image, then they are considered equal for the dimension that this marker image represents, and the comparison outcome is determined by the following dimensions. Hence equivalence groups are now formed at the flat regions of m (Fig. 4.21). The "flattening" process may be achieved using a variety of filters such as large median filters, alternating sequential filters [252], morphological levelings [174], etc. As an application example, let us return to the LSH colour space, and consider once more a lexicographical ordering of type $L \to S$. As such the luminance component would determine the outcome of the vast majority of comparisons. Shifting priority to the second dimension (i. e. saturation) can be achieved using for instance the quantisation example of the previous section. However, since edge related information is represented in its majority at

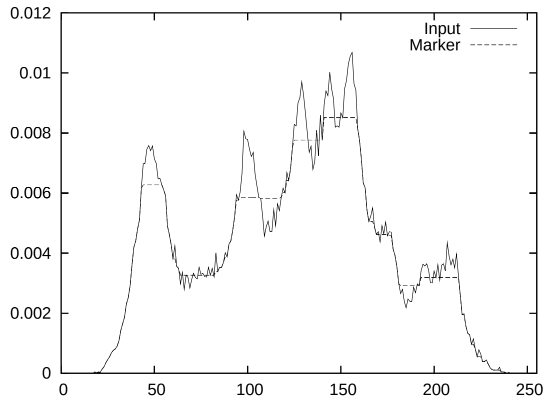


Figure 4.21: Example of a uni-dimensional signal and its marker, obtained with an alternating sequential filter using line shaped SEs of length 5 pixels.

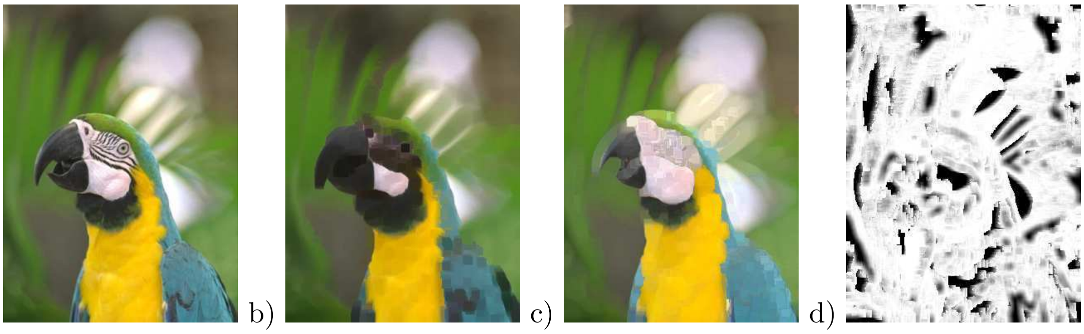


Figure 4.22: The original image (a), its erosion (b) and dilation (c) results with a square shaped SE of size 9×9 pixels, using the marker based ordering of Eq. (4.3.8) along as marker its morphological levelling. The marker of the levelling has been obtained by means of alternating sequential filters using the same SE. And rightmost, is the priority map (d).

the luminance component, it is obviously pertinent to employ this component at object edges and borders, while using saturation for homogeneous or flat image regions. This relatively constrained situation can be resolved with Eq. (4.3.8) as shown in Fig. 4.22.

More precisely, the priority shift towards saturation can be effectively limited to only the relatively homogeneous regions of the input, by flattening them using a morphological levelling applied to the luminance component. The "priority map" shown in Fig. 4.22d asserts this statement. Specifically, this map associates to each pixel the percentage of comparisons within the SE that have been determined using the first dimension (i. e. luminance). Hence dark areas correspond to pixels were the outcome of vector comparisons has been determined using the saturation component, whereas bright areas represent pixels were luminance, the first vector dimension, has been used to order the vectors.

Consequently, although the creation of equivalence groups is image specific, and needs to be realised independently for each image to be processed, marker based ordering provides nonetheless a means of avoiding the artificial edge related pitfalls of the aforementioned approaches, by taking into account the spatial relationships of pixels, thus making it possible to accommodate topological restrictions during ordering.

4.3.3 α-trimmed lexicographical extrema

In this section, we follow a rather unusual approach. Instead of considering the problem of colour/vector ordering as a means to compute colour extrema, that are to be used for defining the dilation and erosion operators, we concentrate directly on the computation of colour extrema, given a set of colours. Thus we can exploit the distribution of the vector/colour set in the multi-dimensional space, in order to better control the lexicographical comparison cascade. Naturally, since there is no underlying ordering, there is no complete lattice structure, hence the resulting operators are theoretically invalid. Nevertheless, as it will be shown in Sec. 4.3.4, they still have a certain practical interest.

Definitions

The α-trimming principle has long been used in filters such as the α-trimmed mean filter (αMF) and its variants [201] against impulsive noise. Given a vector $\mathbf{v} \in \mathbb{R}^n$, containing the sorted scalar pixels under the filtering window, the underlying idea of α-trimming consists in computing their mean by ignoring the 2α extreme:

$$\alpha\mathrm{MF}(\mathbf{v}) = \frac{1}{n - 2\alpha} \sum_{i=\alpha+1}^{n-\alpha} v_i \tag{4.3.9}$$

where $\alpha \in [0, n/2]$. Similarly, in the case of multidimensional vectors we can apply a likewise principle to each dimension in an iterative mode. Specifically, in the case of the maximum, starting from the first dimension, we can sort all k vectors according to this dimension, and then keep the $\lceil \alpha \times k \rceil$ greatest. By repeating this simple process for each dimension, at each step the initial set of vectors will get smaller. In the eventual case where more than one vector remain at the end of this procedure, the last dimension is used for determining the sought extremum. A more formal description for computing the maximum based on this procedure, is given in Algorithm 2.

Algorithm 2 The α-trimmed lexicographical maximum computation algorithm

```
Input: a set V = {v_j} of k, n-dimensional vectors; α ∈ ]0, 1]
Output: max V
for i ← 1 to n − 1 do
   Sort in increasing order the vectors of V with respect to their i^th dimension
   k ← ⌈α × k⌉
   V ← the greatest k vectors in V as well as those equal to the k^th vector
   if (|V| = 1) then
      return v ∈ V
   end if
   i ← i + 1
end for
return the greatest vector within V with respect to the n^th dimension
```

As to the minimum, it can be obtained in a likewise mode by simply sorting in a decreasing order. Consequently, the resulting extrema can be used in order to define the operators in Eqs (3.2.3) and (3.2.4) for multivariate pixels. An illustration of this approach on a three dimensional space $D1 \times D2 \times D3$ is given in Fig. 4.23. Moreover, the advantage of a collective extremum calculation comes however with an increased computational burden. To explain, assuming an optimal sorting procedure is used, in the worst scenario, where all n dimensions would need to be sorted, with k vectors the complexity would be in the order of $O(n \times k \times \log k)$. On the other hand, with the standard lexicographical ordering and the same scenario, the complexity would be $O(n \times k)$. Note that as k represents in practice the size of the SE in use, its value is relatively small.

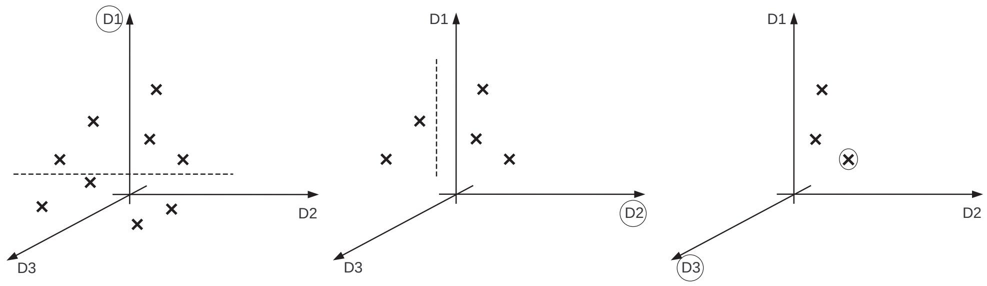


Figure 4.23: A set of vectors in a three dimensional space $D1 \times D2 \times D3$, and the three iterations of the α-trimmed lexicographical maximum computation with $\alpha = 0.5$ (left to right). According to the proposed approach, the maximum is the greatest with respect to the third dimension of the remaining three vectors.

Setting α

The extrema obtained in this way, depend of course directly on the value of α. As a matter of fact, with an α approaching zero, from each dimension i, only the extreme (i. e. maximum or minimum) vector is kept along with those equal to it with respect to the i^{th} dimension. In other words the procedure becomes identical to the standard lexicographical ordering. On the other hand, when α approaches one, priority is shifted gradually to the last dimension, and almost all initial vectors reach the last dimension where the comparison is finally decided.

This transition of priority is illustrated in Fig. 4.24. When keeping only a single value at each dimension, identically to a standard lexicographical ordering, the first dimension decides the outcome of the majority of comparisons. The final decision is shifted rapidly to the last dimension by slightly increasing the number of kept vectors. The main difference with respect to other lexicographical ordering approaches, is the collective extremum calculation, conversely to a binary one. Specifically, even if the final extremum choice is made in the last dimension, unless $\alpha = 1.0$, all previous dimensions contribute to this choice by trimming the set of vectors accordingly. Additionally, more complicated priority relations among the available channels can be established by means of different α values for each vector dimension.

In practice however, it is often necessary to set these arguments in an unsupervised way. Here we introduce a simple parameter setting model based on the standard deviation (σ) of each dimension. More precisely, if the data are relatively concentrated with respect to their i^{th} dimension, or in other words if this image channel does not contain much of the total variational information, we consider using a large α to be more pertinent, thus decreasing the influence of this dimension by carrying the majority of the input to the next dimension with minor trimming. On the other hand, if the data are highly dispersed with respect to

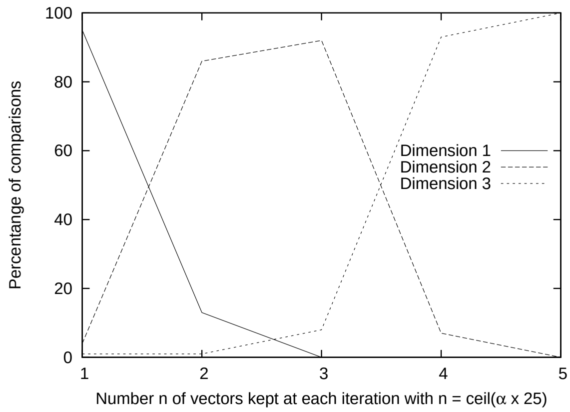


Figure 4.24: The plot of comparisons decided by each dimension of the image Lenna (Fig. 4.2.2 left) in the RGB colour space, during a dilation with a square shaped SE of size 5×5 pixels, for various numbers of vectors kept at each iteration.

the other dimensions, meaning that this channel represents relatively important variational information, a small α would be used leading to major trimming. Given n dimensions, one way of obtaining the corresponding α_i value of dimension i would be:

$$\forall\, i \in \{1, \ldots, n\}\,, \;\; \alpha_i = 1 - \frac{\sigma_i}{\sum_{j=1}^{n} \sigma_j} \tag{4.3.10}$$

where σ_j denotes the standard deviation of dimension j. A more effective way of employing this principle would be to first apply a principal components transformation on the image data and then use the variances of each new dimension to this end.

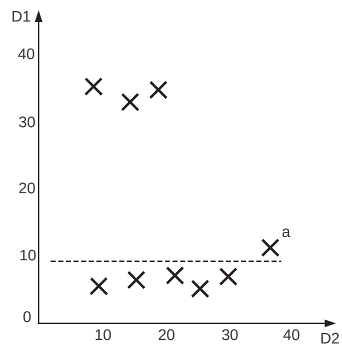


Figure 4.25: Example of problematic vector distribution. The vector **a** is chosen as maximum with $\alpha = 0.4$, while being positioned at a far distance withg respect to the upper vector group.

Variations

In this section, we introduce some variations to Algorithm 2. Using the α argument as a cardinality limit at each iteration of the algorithm can lead to unexpected situations, such as the one shown in Fig. 4.25. Indeed, by selecting the vectors to be used in the next stage of the cascade based only on the cardinality of the vector set (i. e. $\lceil \alpha \times k \rceil$), in other words irrespectively to their relative distances, one may end up with extremely scattered vector clusters. This situation can be easily countered by using α as a distance measurement. Specifically, instead of keeping the $\lceil \alpha \times k \rceil$ greatest vectors with respect to the i^{th} dimension, one can keep only the vectors whose i^{th} dimension is at most at distance α with respect to the greatest vector of the same dimension. Details are given in Algorithm 3.

Algorithm 3 The α-trimmed lexicographical maximum computation algorithm based on distances.

Input: a set $V = \{v_j\}$ of k, n-dimensional vectors; $\alpha \in]0, 1]$
Output: $\max V$
for $i \leftarrow 1$ to $n - 1$ **do**
  Calculate v_{min} and v_{max}, the smallest and largest vectors
    with respect to their i^{th} dimension
  $V \leftarrow$ the set of vectors v_j satisfying $d_i(v_{max}, v_j) \leq \alpha \times d_i(v_{max}, v_{min})$,
    where d_i is the scalar distance of their i^{th} dimension
  if ($|V| = 1$) **then**
    return $v \in V$
  end if
  $i \leftarrow i + 1$
end for
return the greatest vector within V with respect to the n^{th} dimension

---

As previously mentioned, a more important "disadvantage" however of α-trimmed lexicographical extrema is the lack of an underlying binary ordering relation. Nevertheless, given any extremum computation method, it is possible to construct from it an ordering. In particular, considering a discrete multi-dimensional space, one can employ the collective extremum computation method at hand, in order to calculate the maximum or minimum of this space, and then repeat for the remaining points, up until the entire space is ordered. The formal procedure is described in Algorithm 4.

Algorithm 4 The ordering of a multi-dimensional space using a collective minimum computation method (CollectiveMinimum).

Input: a set $V := \{v_i\}_{1 \leq i \leq k}$ containing all the vectors of a discrete multi-dimensional space
Output: the ordered sequence $V' := \left\{ v'_j \in V \,|\, \forall m, n \in \{1, \ldots, k\}, m \leq n, v'_m \leq v'_n \right\}$ containing the vectors of V
for $j \leftarrow 1$ to k **do**
  $v'_j \leftarrow$ CollectiveMinimum $\{V \setminus \{v'_i \,|\, i \in \{1, \ldots k\}, i < j\}\}$
  $j \leftarrow j + 1$
end for
return $V' \leftarrow \left\{ v'_j \right\}_{1 \leq j \leq k}$

Moreover, since α-trimmed lexicographical extrema are unique, this approach leads to a total ordering. The exact algebraic properties of the result however depend strongly

Figure 4.26: From left to right, the original image Lenna, its dilation and erosion, both with a square shaped SE of size 7×7 using α-trimmed lexicographical extrema with α set by Eq. (4.3.10) in the LSH colour space.

both on the extremum computation method used (e. g. α-trimmed lexicographical, cumulative distances, etc) as well as on the extremum employed (i. e. minimum or maximum) during its construction.

Case study: colour pseudo-morphology

The so far discussed extrema computation mechanisms have been independent of the underlying multi-dimensional space. In this section, we study the specific case of colour images, and illustrate the application of colour pseudo morphological operators in this context.

As the human vision system is largely known to attribute greater importance to brightness variations with respect to colour, it is decided to set the lexicographical comparison order as $(L \rightarrow S \rightarrow H)$ where all components are normalised to $[0, 1]$, and the hue is processed according to Eq. (3.4.6), with the reference hue being $h_0 = 0.0$. Image specific reference points have been additionally tested, however as the hue occupies the last position of the lexicographical cascade, their impact on the end results has been observed to be negligible in the context of these experiments. Consequently, the dilation and erosion operators can be defined by calculating the necessary colour extrema according to Algorithm 2.

Fig. 4.26 illustrates the application of the two operators on the Lenna image. Additionally, the effect of α-trimmed extrema can be better observed from Figs. 4.27 and 4.28, which illustrate respectively the average values of each image channel for dilations computed based on the proposed trimmed extrema (Algorithm 2) and α-modulus lexicographical ordering, Eq. (4.3.2), using varying values for α. Note that erosions lead to symmetrical plots with respect to the horizontal axis.

As far as luminance is concerned (Fig. 4.27a), since small values of α render the ordering an approximation of its standard lexicographical counterpart, one can clearly observe a monotonic decrease in the average value, after of course the rapid increase from its original input value for $\alpha = 1/25$, conversely to the case of α-modulus lexicographical ordering which exhibits a more unstable behaviour (Fig. 4.28a). By increasing α and hence the number of vectors kept from each dimension, comparison priority is rapidly shifted towards the second channel, i. e. saturation (Fig. 4.27b). Whereas the hue (Fig. 4.27c), as the value of α approaches one, determines almost exclusively the outcome of vector comparisons, thus modifying the image's average distance from the reference hue accordingly.

As to the behaviour of α-modulus lexicographical ordering, one can remark the lower

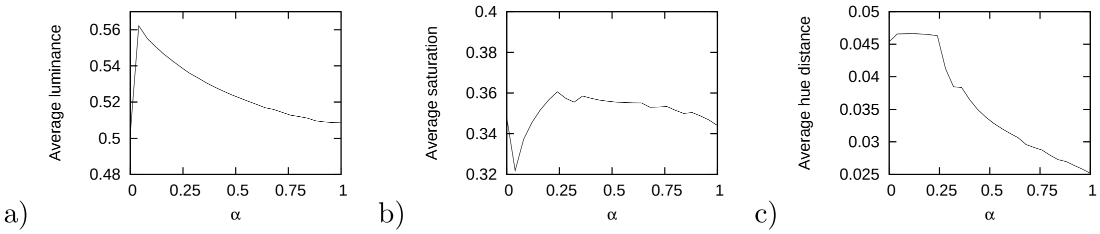


Figure 4.27: The plots of average luminance (a), average saturation (b) and average hue distance (c) from $h_0 = 0.0$, computed with Eq. (3.4.5), of the Lenna image, during dilations with a square shaped SE of size 5×5 pixels using the trimmed extrema of Algorithm 2 combined with varying values of α.

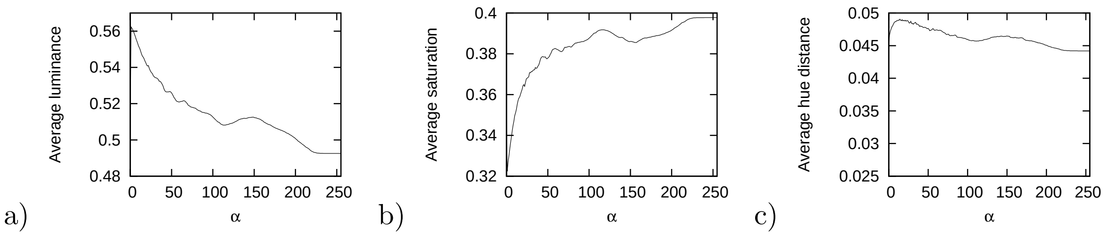


Figure 4.28: The plots of average luminance (a), average saturation (b) and average hue distance (c) from $h_0 = 0.0$, computed with Eq. (3.4.5), of the Lenna image, during dilations with a square shaped SE of size 5×5 pixels using the α-modulus lexicographical ordering, Eq. (4.3.2), combined with varying values of α.

in average luminance levels (Fig. 4.28a), as well as the greater distances to the reference hue (Fig. 4.28c), while saturation is much more privileged for relatively high values of α (Fig. 4.28b). This result is obviously due to the fact that the division taking place within the first dimension of vectors favours almost exclusively the second. Furthermore in the case of α-modulus lexicographical ordering, also worth noting is the decrease of luminance *below* the initial level of the input for higher α values, hence leading to a darker image after dilation. As these results are provided as mere indications and are dependent on the image at hand, we further proceed to realise a series of comparative tests among the different approaches in order to better evaluate their performances.

4.3.4 Applications

In this section, the previously presented three lexicographical ordering based approaches are tested through noise reduction and texture classification applications, along with colour images in the LSH space. In order to carry out the various quantisation steps the luminance channel is manipulated with integer precision in $[0, 255]$. In particular, the extrema employed for defining the basic morphological operators, are computed using only the luminance channel (Lum), only saturation (Sat), only hue (Hue), lexicographical ordering, Eq. (3.2.6) (Lex), α-modulus lexicographical ($\alpha = 10$), Eq. (4.3.2) (α-modLex), quantisation based α-lexicographical ($\alpha = 10$), Eq. (4.3.7) (QuantaLex) with $\alpha = 10$, the marker based lexicographical ordering, Eq. (4.3.8) (MarkerLex), where the marker image is obtained by applying a leveling (detailed in Sec. 6.2.2) along with a leveling marker image provided by median filtering with a square-shaped SE of size 7×7, the α-trimmed lexicographical extrema (α-trimmedLex), according to Algorithm 2 and $\alpha = 0.45$, as well as its variations: the one using adaptive α values, Eq. (4.3.10) (α-trimmedLex-adaptive), the total ordering resulting from the application of Algorithm 4, with $\alpha = 0.45$ (α-trimmedLex-ordering), and finally the version based on distance computations, given in Algorithm 3 (α-trimmedLex-distance) with $\alpha = 0.3$. The comparisons are realised in the order $L \to S \to H$, and the hue component is processed according to the principle of weighted multiple references, as in Eq. (4.2.8). Besides, for both applications we employ basically the same experimental setups as in Secs. 3.4.1 and 3.4.2.

	$\rho = 0.0$	$\rho = 0.95$
Lum	2.14	0.89
Sat	2.76	2.94
Hue	2.64	3.04
Lex	2.14	0.89
α-modLex	2.16	0.91
QuantaLex	2.13	0.88
MarkerLex	2.45	1.17
α-trimmedLex	**1.71**	**0.84**
α-trimmedLex-adaptive	1.72	0.85
α-trimmedLex-ordering	3.91	3.04
α-trimmedLex-distance	2.25	1.01

Table 4.2: $100 \times NMSE$ error values against uncorrelated ($\rho = 0.0$) and correlated ($\rho = 0.95$) Gaussian noise ($\sigma = 0.125$)

Noise reduction

Table 4.2 shows the error rates that have been obtained against uncorrelated and correlated Gaussian noise, using the aforementioned ordering options for constructing the OCCO filter. As far as the case of uncorrelated noise is concerned, we can observe the superiority of luminance over the other two dimensions, hence justifying our choice to set luminance at the first position of the lexicographical comparisons. Furthermore, one can easily remark the extremely similar values obtained by using only the luminance channel for comparisons and the standard lexicographical ordering. A result which asserts the highly asymmetrical priority attributed to the first vector component. At the first group of lexicographical variations, the transition of priority to the saturation channel by means of α-modLex does not result in an improvement, leading us to assume that the luminance channel is more pertinent for this task. However, the proposed QuantaLex approach disproves this assumption, though only slightly superior than its lexicographical counterpart, by exploiting the inner relationships among luminance and saturation. As to MarkerLex, it is obviously not suitable for the task of noise reduction, since its highly smoothed marker image prevents it from accessing the fine differences among the corrupt pixels of the luminance channel.

Two of the four trimming based extrema on the other hand, clearly outperform their counterparts. Despite resulting in pseudo-morphological operators, collective extrema computation is presumed to aid their performance. Furthermore, although the empirically set α value provides the best results, the adaptively set dimension specific argument (α-trimmedLex-adaptive) exhibits only slightly superior error levels. α-trimmedLex-ordering as well as α-trimmedLex-distance however have been disappointing, clearly showing that further work is necessary on these approaches.

As to the case of correlated Gaussian noise, overall equivalent performances have been obtained. Despite using an empirically set optimum α value, for α-trimmedLex, it hardly surpasses its counterparts. A possible explanation for this result is the presence of correlated noise in its majority, within the luminance dimension, hence rendering it relatively redundant to take into account saturation and hue. Additionally, one can also observe the satisfying results obtained once more with the adaptively set argument of α-trimmedLex-adaptive, asserting the robustness of the proposed model.

	Accuracy (%)
Lum	73.82
Sat	66.91
Hue	55.29
Lex	74.26
α-modLex	76.62
QuantaLex	77.17
MarkerLex	79.86
α-trimmedLex	**81.47**
α-trimmedLex-adaptive	79.89
α-trimmedLex-ordering	54.35
α-trimmedLex-distance	73.74

Table 4.3: Classification accuracies for the textures of Outex13, using vector erosion based covariance.

Texture classification

The results obtained from the texture classification application are given in Table 4.3. The overall performances appear to be relatively close. Specifically, the pertinence of luminance over both saturation and hue is once more observed, showing that colour information is in this case just an auxiliary component. As expected, Lum and Lex exhibit almost identical performances, a result showing the level of priority attributed to the first component, ignoring largely the remaining two dimensions. A minor improvement over Lex is obtained through α-modLex, as it allows saturation to participate at a slightly higher degree in the process of computing covariance. However, hue remains largely unexploited. With the use of α-trimmed extrema, the chrominance dimensions increase their contribution to the final result. As a matter of fact, the empirically set α-trimmedLex leads to an overall best, while the unsupervised model once more provides a relatively sufficient approximation. Nevertheless, no improvements are observed with the last two variations.

Moreover, QuantaLex leads again to an improvement over α-modLex, as it controls the transitions among the image channels in a finer way, the difference being however almost negligible. Contrary to noise reduction, the marker based approach improves its performance substantially and outperforms α-modLex. Apparently, the use of saturation with relatively "flat" regions, and that of luminance with transitions within the image, lead to more pertinent descriptors by better combining colour and textural information.

4.4 Colour hit-or-miss transform

The *hit-or-miss transform* (HMT) is a powerful morphological tool, that was among the initial morphological transforms developed by Matheron and Serra [171, 240]. It constitutes the morphological approach to pattern matching. Its initial definition for binary images has been widely used since then, with the purpose of shape recognition [24, 28], while multiple theoretical extensions have been proposed in order to improve its performance [28, 251, 301]. The extension of this transform to grey-level images however has not been straightforward, since it is neither increasing nor decreasing, hence leading to multiple definitions, such as those of Ronse [226], Soille [252] and Barat et al. [20]. Only recently a unified theoretical framework has been proposed for grey-level HMTs by Naegel et al. [186].

As far as colour images are concerned, the potential of the HMT has been left largely unexplored [292], with the exception of Ref. [286], which presents a novel approach for a multivariate HMT based however on a marginal strategy. The reason of this situation, besides the aforementioned ambiguities of its extension to grey-level data, lies additionally in the complications inherent to the application of morphological operators to multivariate images, that have been elaborated in Sec. 3.2. Despite these difficulties, a vector HMT has an important potential, that can be exploited for instance with the purpose of object detection from remote sensing data, or colour object recognition.

Motivated by this application potential, we propose a vector HMT (VHMT) definition, capable of detecting objects from colour images, based on colour templates. Its main advantage with respect to a marginal or component-wise definition, lies in its capacity to be further configured by means of the vector ordering choice, hence allowing the introduction of a priori information concerning the sought object's inter-channel presence. Furthermore, since the HMT is inherently sensitive to image variations, a couple of approaches aiming to counter this drawback, already used with binary and grey-level images [20, 59, 251], are studied in the case of colour data. Specifically, rank order filters as well as "synthetic" structuring functions are discussed.

After presenting the case of binary and grey-level HMT in Sec. 4.4.1, the vector case is elaborated in Sec. 4.4.2, where the VHMT definition is given. Then Sec. 4.4.3 presents the adaptation to the vector case of two known techniques for increasing the robustness of the HMT.

4.4.1 Binary and grey-level HMT

In this section, we will review briefly the binary and grey-level approaches to the HMT. The initial binary definition of the HMT [171, 240], consists in searching in the input binary image $X \in \mathcal{P}(\mathcal{E})$, $\mathcal{E} = \mathbb{R}^d$ or $\mathbb{Z}^d$, a template described by a couple of structuring elements (SE), $A, B \in \mathcal{P}(\mathcal{E})$ (Fig. 4.29). Specifically, it attempts to match A *(foreground SE)* within the image (i. e. "*a hit*") while also matching B *(background SE)* in its background $X^c = \mathcal{E} \setminus X$ (i. e. "*a miss*"):

$$X \circledast (A, B) = (X \ominus A) \cap (X^c \ominus B) \tag{4.4.1}$$

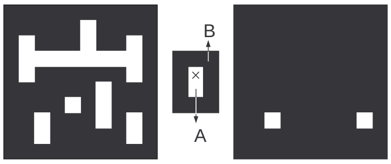


Figure 4.29: From left to right, the input binary image X, the sought template (A, B), and the output of Eq. (4.4.1). The "×" within A marks the center of the SE.

where $\ominus$ denotes the binary erosion operator, and it is assumed that $A \cap B = \emptyset$. However, since Eq. (4.4.1) uses both X and X^c, its extension to grey-level images has not been straightforward.

As a matter of fact, this extension was not realised until after the HMT was expressed in terms of an *interval operator* η, through the work of Heijmans and Serra [109]. For $A, B \in \mathcal{P}(\mathcal{E})$, $A \subseteq B$:

$$\eta_{[A,B]}(X) = \{p \in \mathcal{E} \mid A_p \subseteq X \subseteq B_p\} = X \circledast (A, B^c) \tag{4.4.2}$$

where $A_p = \{x + p \mid x \in A\}$ denotes the translate of A by p. Through this formulation, the grey-level HMT of Ronse [226] and subsequently of Soille [252] have been defined, and along with others [20, 237, 134, 218], they have been recently unified into a common theoretical framework for grey-level interval operators [186].

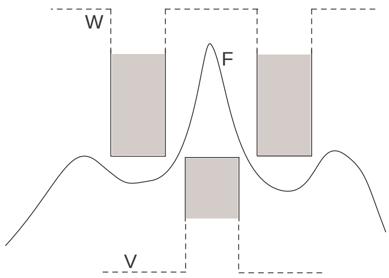


Figure 4.30: The integral interval operator, Eq. (4.4.7).

More precisely, in the case of grey-level images, an image $F \in \mathcal{T}^{\mathcal{E}}$, ($\mathcal{T}$ being a complete lattice, usually $\overline{\mathbb{R}}$ or $\overline{\mathbb{Z}}$), besides a SE, can also interact with a structuring function (SF) $V \in \mathcal{T}^{\mathcal{E}}$. Consequently, the sought template is translated not only horizontally (by a point $p \in \mathcal{E}$), but vertically as well (by a finite grey-level $t \in \mathcal{T}$) (Fig. 4.30) in an attempt to detect the positions where it fits. Specifically, a translation by a couple (p, t) is $V_{(p,t)} : x \to V(x - p) + t$. According to the unified theory for grey-level HMT [186], such an operator is decomposed into two stages:

- the *fitting*, where the locations fitting the given structuring functions, describing the sought template,. are computed,
- and *valuation*, where the resulting image containing the previously detected locations is constructed.

To explain, there are two types of fittings. Given a grey-level image F, along with a couple of structuring functions $V,\ W,\ (V \leq W)$ describing the sought template, there is first:

$$H_{V,W}(F) = \left\{(p,t) \in \mathcal{E} \times \mathcal{T} \mid V_{(p,t)} \leq F \leq W_{(p,t)}\right\} \tag{4.4.3}$$

the inequality of which according to Ref. [186] is equivalent to:

$$\varepsilon_V(F)(p) \leq t \leq \delta_{W^*}(F)(p) \tag{4.4.4}$$

where $W^* : x \rightarrow -W(-x)$, is the *dual* of W and ε, δ represent respectively the grey-level erosion and dilation with a SF. The other fitting is:

$$K_{V,W}(F) = \left\{(p,t) \in \mathcal{E} \times \mathcal{T} \mid V_{(p,t)} \leq F \ll W_{(p,t)}\right\} \tag{4.4.5}$$

where $F \ll W$ means that there is some $h > 0$ such that for every $p \in \mathcal{E}$ we have $F(p) \leq W(p) - h$. Besides, the inequality of the second fitting once more according to Ref. [186] is equivalent to:

$$\varepsilon_V(F)(p) \leq t < \delta_{W^*}(F)(p) \tag{4.4.6}$$

Likewise, there are two[1] types of valuations. First there is *supremal* valuation, which for every point p of fit couple (p,t), takes the supremum of t, and then there is *integral* valuation which instead for every point p of fit couple (p,t), uses the length of the interval of t for which the couples (p,t) fit. Moreover, Soille's grey-level HMT [252] employs the fitting K of Eq. (4.4.5) along with an integral valuation, which leads to the *integral interval operator* η^I (Fig. 4.30):

$$\eta^I_{[V,W]}(F)(p) = \max\left\{\varepsilon_V(F)(p) - \delta_{W^*}(F)(p), 0\right\} \tag{4.4.7}$$

It should be noted that the reason for choosing the fitting K instead of H is that it produces a semi-open interval $[\varepsilon_V(F)(p), \delta_{W^*}(F)(p)[$ and not a closed one. Thus in the discrete case the interval length formulation corresponds indeed to that of Eq. (4.4.7). Ronse's grey-level HMT on the other hand, uses the fitting H of Eq. (4.4.3) combined with a supremal valuation, which leads to the *supremal interval operator* η^S:

$$\eta^S_{[V,W]}(F)(p) = \begin{cases} \varepsilon_V(F)(p) & \text{if } \varepsilon_V(F)(p) \geq \delta_{W^*}(F)(p), \\ \perp & \text{otherwise} \end{cases} \tag{4.4.8}$$

where $\perp$ is the least element of $\mathcal{T}$, $-\infty$ if $\mathcal{T} = \overline{\mathbb{R}}$. In the next section, the grey-level integral operators of Eqs. (4.4.7) and (4.4.8) will be extended to multivariate images.

4.4.2 Vector HMT

Given Eqs. (4.4.7) and (4.4.8) in combination with multivariate image data, the obvious way of searching for a multivariate template would be to apply them independently to each image channel. This principle is endorsed in Ref. [292], where binary multivariate images are considered. A more original approach using channel specific thresholds is presented in Ref. [286], which however is still based on a marginal strategy. In short, these approaches ignore any eventual correlation among image channels, thus the alternative consists in defining multivariate morphological operators.

[1] In fact there is also *binary* valuation, which consists in taking the set of points $p \in \mathcal{E}$ for which there is at least one $t \in \mathcal{T}$ such that (p,t) fits.

In the multivariate case, the pixel values of images are now in $\mathcal{T} = \overline{\mathbb{R}}^n$ or $\overline{\mathbb{Z}}^n, n > 1$. The initial step consists in defining the erosion and dilation operators for multivariate images in combination with multivariate SF. More precisely, these operators are once more based on horizontal translations (by a point $p \in \mathcal{E}$) as well as on vertical ones (by a finite pixel value $\mathbf{t} \in \mathcal{T}$) as in the grey-level case, the difference is however that pixel values are now multi-dimensional; in particular, given a multivariate image $\mathbf{F} : \mathcal{E} \to \mathcal{T}$:

$$\forall\, (p, \mathbf{t}) \in \mathcal{E} \times \mathcal{T},\;\; \mathbf{F}_{(p,\mathbf{t})}(x) = \mathbf{F}(p - x) + \mathbf{t} \tag{4.4.9}$$

Furthermore, according to the fundamental Refs. [110, 112] of Heijmans and Ronse, translations need to be complete lattice automorphisms (i. e. bijections $\mathcal{T} \to \mathcal{T}$ that preserve order, and whose inverse also preserve order). Consequently, the vector ordering ($\leq_v$) from which the complete lattice is derived, must be translation invariant. In other words:

$$\forall\, \mathbf{w}, \mathbf{w}', \mathbf{t} \in \mathcal{T},\;\; \mathbf{w} \leq_v \mathbf{w}' \Leftrightarrow \mathbf{w} + \mathbf{t} \leq_v \mathbf{w}' + \mathbf{t} \tag{4.4.10}$$

Thus we can give the definition of the erosion and dilation respectively of a multivariate image $\mathbf{F}$ by a multivariate SF $\mathbf{B}$:

$$\boldsymbol{\varepsilon}_{\mathbf{B}}(\mathbf{F})(p) = \inf_{v\; x \in \mathrm{supp}(\mathbf{B})} \{\mathbf{F}(p + x) - \mathbf{B}(x)\} \tag{4.4.11}$$

$$\boldsymbol{\delta}_{\mathbf{B}}(\mathbf{F})(p) = \sup_{v\; x \in \mathrm{supp}(\mathbf{B})} \{\mathbf{F}(p - x) + \mathbf{B}(x)\} \tag{4.4.12}$$

where $\mathrm{supp}(\mathbf{B}) = \{p \in \mathcal{E} \mid \mathbf{B}(p) > \perp\}$. Hence, these formulations form an adjunction as demanded by Refs. [110, 112] and besides, with a flat SE (i. e. $\forall x$, $\mathbf{B}(x) = \mathbf{0}$) they are reduced to the flat multivariate erosion and dilation formulations given respectively in Eqs. (3.2.3) and (3.2.4). Furthermore, thanks to Eq. (4.4.10), both fitting equivalences between Eqs. (4.4.3) and (4.4.4) as well as between Eqs. (4.4.5) and (4.4.6) become directly extendable to this case by replacing the grey-level operators with their multivariate counterparts. As to valuation, the same options as before are available, however the supremum is of course now computed among vectors through the vector ordering in use. In the case of integral valuation, a vector distance now can be used in order to measure the distance among the vectors that have fit. Consequently one can express the multivariate versions of the integral and supremal interval operators respectively as follows:

$$\boldsymbol{\eta}^I_{[\mathbf{V},\mathbf{W}]}(\mathbf{F})(p) = \begin{cases} \|\boldsymbol{\varepsilon}_{\mathbf{V}}(\mathbf{F})(p) - \boldsymbol{\delta}_{\mathbf{W}^*}(\mathbf{F})(p))\| & \text{if } \boldsymbol{\varepsilon}_{\mathbf{V}}(\mathbf{F})(p) >_v \boldsymbol{\delta}_{\mathbf{W}^*}(\mathbf{F})(p) \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.13}$$

$$\boldsymbol{\eta}^S_{[\mathbf{V},\mathbf{W}]}(\mathbf{F})(p) = \begin{cases} \boldsymbol{\varepsilon}_{\mathbf{V}}(\mathbf{F})(p) & \text{if } \boldsymbol{\varepsilon}_{\mathbf{V}}(\mathbf{F})(p) \geq_v \boldsymbol{\delta}_{\mathbf{W}^*}(\mathbf{F})(p) \\ \perp & \text{otherwise.} \end{cases} \tag{4.4.14}$$

$$\tag{4.4.15}$$

where $\mathbf{V} \leq_v \mathbf{W}$. As far as $\boldsymbol{\eta}^I$ is concerned, it provides a non-zero output at positions where $\mathbf{V} \leq_v \mathbf{F} \ll_v \mathbf{W}$ according to the ordering in use. It should also be noted that the grey-level valuation choice by means of the Euclidean norm ($\|\cdot\|$) is arbitrary, and a multi-dimensional valuation is of course possible. As to $\boldsymbol{\eta}^S$, it produces a non-zero output at positions where $\mathbf{V} \leq_v \mathbf{F} \leq_v \mathbf{W}$.

Considering now the need for a translation preserving vector ordering, the lexicographical option represents an adequate choice for this task. Moreover, the chosen ordering directly affects the behaviour of VHMT. For instance, in the case of lexicographical ordering, which is known for its tendency of prioritising the first vector component, this property can be observed in the detection process of VHMT. In particular, during the fitting stage where the erosion and dilation outputs are computed, since it is the first vector component that decides the outcome of the majority of lexicographical comparisons, fitting the first channel of the vector structuring function becomes more important with respect to the rest. This example is illustrated in Fig. 4.31, where although only $V_1 \leq F_1$, according to the lexicographical ordering, $\mathbf{V} <_L \mathbf{F}$. This property for example would allow for a prioritised detection, where a colour template is searched with more emphasis on its brightness than its saturation.

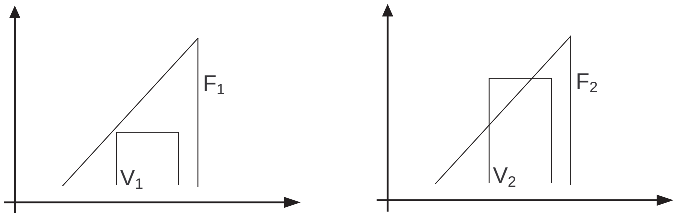


Figure 4.31: A two-channel image $\mathbf{F} = (F_1, F_2)$ and a vector structuring function $\mathbf{V} = (V_1, V_2)$.

A more practical example of VHMT is given in Fig. 4.32, where the yellow sign of the middle is sought using a lexicographical ordering in the LSH space of type $(L \to S)$. More precisely, the SF positioned under the object ($\mathbf{V}$) is formed by decreasing the pixel values of the template by a fixed amount (e. g. 3, if pixel values are in $[0, 255]$), whereas the background SF ($\mathbf{W}$) is formed by increasing it. Hence, the operator looks for all objects that fit between the upper and lower SF based on the lexicographical principle. In this particular case, as the hue does not participate in the ordering, it detects the left sign despite its different hue value, while it misses the right sign, even though its only difference from the template are a few white points; a result that asserts the sensitivity of the operator.

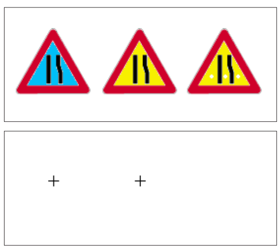


Figure 4.32: In the first row are three images, of which the middle is the sought pattern. The second row shows the locations where it was detected by Eq. (4.4.13) based on a lexicographical ordering of luminance and saturation.

4.4.3 An approximative VHMT

As the HMT attempts to perform an exact match of the given pattern, it becomes sensitive to noise, occlusions, and even to the slightest variations of the shape of the object to detect.

Consequently a series of approaches have been implemented for the binary and grey-level versions of the operator, with the purpose of countering this drawback and increasing its practical interest. Among others, rank order filters [226, 251] and "synthetic" SE [20, 59] can be mentioned. This section concentrates on these two methods and studies their use in the case of multivariate images in combination with VHMT.

Rank order based VHMT

A rank order filter of k^{th} rank is a transformation, where given a grey-level image F and a SE B, the result becomes:

$$(F\Box_k B)(p) = k^{th} \text{ largest of } F(p+x),\ x \in B \tag{4.4.16}$$

with $k \in \{1, \ldots, |B|\}$. Obviously, $F\Box_1 \check{B}$ and $F\Box_{|B|}B$ with $\check{B} = \{-x|x \in B\}$ the reflection of B, correspond respectively to the dilation and erosion of F by B. Moreover, always in the context of grey-level data, a rank order filter of k^{th} rank, is equivalent to the supremum of erosions using all possible SE with k points, and respectively to the infimum of dilations using all possible SE with $|B| - k + 1$ points [252]. Due to this property, the binary HMT of Eq. (4.4.1) has been reformulated in the literature, by replacing the erosion in its expression with a rank order filter of rank $k < |B|$, hence making it possible to detect binary templates even in conditions of partial occlusion.

In order to achieve the same additional robustness in the case of a multivariate image $\mathbf{F}$ along with a multivariate SF $\mathbf{B}$ and a vector ordering $\leq_v$, one can redefine the rank order filter of k^{th} rank as follows:

$$\zeta^k_{\mathbf{B}}(\mathbf{F})(p) = k^{th} \text{ largest of } \mathbf{F}(p+x) - \mathbf{B}(x),\ x \in \operatorname{supp}(\mathbf{B}) \tag{4.4.17}$$

$$\theta^k_{\mathbf{B}}(\mathbf{F})(p) = k^{th} \text{ largest of } \mathbf{F}(p-x) + \mathbf{B}(x),\ x \in \operatorname{supp}(\mathbf{B}) \tag{4.4.18}$$

where $k \in \{1, \ldots, |\operatorname{supp}(\mathbf{B})|\}$. Naturally, the vectors are sorted using $\leq_v$. Thus, $\boldsymbol{\varepsilon}_{\mathbf{B}} = \zeta^{|\operatorname{supp}(\mathbf{B})|}_{\mathbf{B}}$ and $\boldsymbol{\delta}_{\mathbf{B}} = \theta^1_{\mathbf{B}}$. Consequently, we can now formulate an approximative VHMT, capable of detecting the sought template $(\mathbf{V}, \mathbf{W})$ even if m and n pixels do not match respectively the foreground and the background:

$$\boldsymbol{\eta}^I_{[\mathbf{V},\mathbf{W}],m,n}(\mathbf{F})(p) = \begin{cases} \left\|\zeta^m_{\mathbf{V}}(\mathbf{F})(p) - \theta^n_{\mathbf{W}^*}(\mathbf{F})(p))\right\| & \text{if } \zeta^m_{\mathbf{V}}(\mathbf{F})(p) >_v \theta^n_{\mathbf{W}^*}(\mathbf{F})(p) \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.19}$$

where $m \in \{1, \ldots, |\operatorname{supp}(V)|\}$ and $n \in \{1, \ldots, |\operatorname{supp}(W)|\}$. It should also be noted that $\boldsymbol{\eta}^I_{[\mathbf{V},\mathbf{W}],|\operatorname{supp}(W)|,1} = \boldsymbol{\eta}^I_{[\mathbf{V},\mathbf{W}]}$. Fig. 4.33 contains an example of the result given by this operator, where the leftmost image is sought under the same conditions as in Fig. 4.32. However, this time even though the right example has a red/brown stripe, it is still succesfully detected. This is due to the use of the 750^{th} rank, a number equal to the amount of different pixels between the two images. Thus the rank based operator can allow a flexibility margin large enough to realise the detection in case of pixel value variations, due to reasons such as noise.

Synthetic structuring functions

Although multivariate rank order filters make it possible to detect partial matches, Eq. (4.4.19) still hardly satisfies practical needs, since the objects corresponding to the sought template

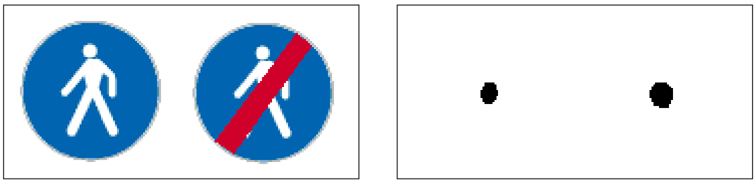


Figure 4.33: On the left is a couple of images, of which the leftmost is the sought pattern, and on the right is the output of $\boldsymbol{\eta}^{I}_{[\mathbf{V},\mathbf{W}],750,750}$, Eq. (4.4.19).

may vary considerably. Consider for instance the case illustrated in Fig. 4.34 (top-left). This situation is of course present in the context of detection from grey-level images as well. One way of countering it, as explained in [20, 59], is to employ a set of example images, from which a common template is formed, or as defined in Ref. [59], "synthetic".

More precisely, the foreground is represented by the minimum and the background by the maximum of the given set of examples $(\{\mathbf{V}_i\}, \{\mathbf{W}_j\})$. Thus, in the multivariate case the same technique may be employed merely by using the chosen vector ordering $\leq_v$:

$$\mathbf{V}(x) = \inf_{i}{}_v \{\mathbf{V}_i(x)\} \tag{4.4.20}$$

$$\mathbf{W}(x) = \sup_{j}{}_v \{\mathbf{W}_j(x)\} \tag{4.4.21}$$

Returning to Fig. 4.34, it suffices to compute the templates corresponding to the images at the top-left by means of this operation, and the VHMT of Eq. (4.4.13) detects both successfully.

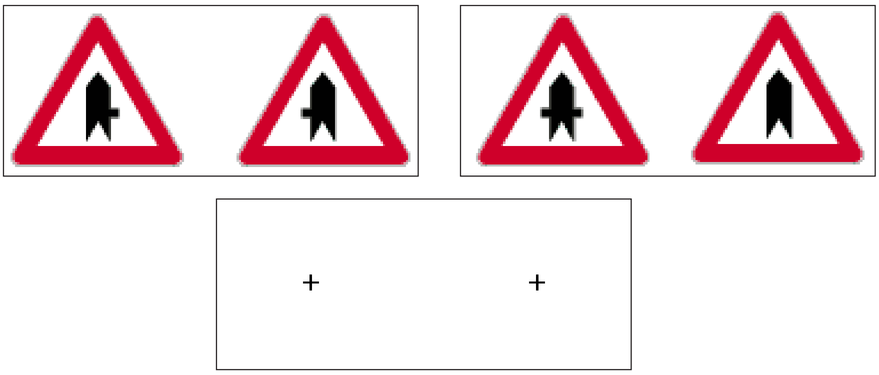


Figure 4.34: In the first row, from left to right the two images to detect, and the corresponding lower and upper synthetic SF computed through Eqs. (4.4.20) and (4.4.21); the second row contains the result of VHMT, Eq. (4.4.13).

4.5 Conclusion

This chapter has aimed to apply to morphology the ordering and colour space choice that has been made in the previous chapter. The combination of lexicographical ordering with phenomenal spaces, and in particular with LSH has already been carried out, however, even though there is no theoretical obstacle for the definition of morphological operators in this way, it has been shown that two issues seriously hinder their practical use. These issues are the processing of the 2π periodical hue value, and the extreme prioritisation of the first vector dimension by lexicographical ordering.

As far as the processing of the hue component is concerned, the most widely used approach consists in manipulating hue values according to their distances from a reference hue value, thus effectively resolving the problem of periodicity. Nevertheless, given the presence of multiple dominant hues in most colour images, we believe this approach to be insufficient in this regard. That is why it has been proposed to use a similar reduction technique, with multiple reference hues, where hue values are ordered according to their distance to the closest reference. Of course appropriate weighting mechanisms are also presented in order to eliminate the effect of relatively unimportant hue groups, as well as a means for obtaining the reference hues in an unsupervised way.

Moreover, the asymmetrically prioritised lexicographical ordering has been shown to lead to the unexploitation of the available image channels. Three different variations of lexicographical ordering have been proposed with the purpose of resolving this issue. The first, quantisation based α-lexicographical ordering, consists of a generalisation of the α-modulus lexicographical ordering. Our method adopts a similar principle, by sub-quantising the vector dimensions according to arbitrary models, hence making it possible for the user to incorporate into the ordering any a priori information available on the channels under consideration. A further effort has been made to model the relations among luminance, saturation and hue, that have been then used in combination with this approach. The second variation of lexicographical ordering, marker based lexicographical ordering, uses an image specific principle, where the first channel is processed by means of grey-level operators, in order to form flat zones, that determine the areas where priority is shifted to the subsequent dimensions. Additionally, a third method for computing α-trimmed lexicographical extrema has been introduced, which also succeeds in manipulating the contribution of each channel to this process through a user defined argument. Nevertheless, as there is no underlying ordering it leads to pseudo-morphological colour operators. Variations of this approach, as well as the aforementioned ones, have been further tested with colour image applications. Even though the α-trimmed extrema based option has led to favorable results, thus asserting some practical interest, in the rest of this thesis we shall focus on the quantisation based variation of lexicographical ordering, Eq. (4.3.7) ($\alpha = 10$). This choice is based on its theoretical validity, as well as capacity to accommodate the channel relations of LSH. It should be further noted, that the proposed ordering variations have been designed with no particular image context in mind, hence being usable in the general case of multivariate image data, e. g. hyperspectral remote sensing images, multispectral astronomical images, etc.

Furthermore, a vector HMT definition is proposed, which allows for the use of vector structuring functions, making it possible to detect multivariate templates within multivariate data such as colour images. The additional parameter of choosing a vector ordering, opens up new combination possibilities, allowing to refine the detection properties of the resulting operator. Two classical techniques for enhancing the flexibility of HMT have also been studied in combination with colour images. Perspectives on this subject would include the study of the effects of the vector ordering choice on VHMT, as well as ways of rendering the resulting operators rotation and scale invariant.

The rest of this thesis will elaborate on the application of colour morphological operators to the problem of general purpose content-based image retrieval, starting with an overview of this field.

Part II

Mathematical morphology applied to content-based colour image retrieval

Chapter 5

Content-based image retrieval

5.1 Introduction

Digital images are everywhere. The constantly improving capabilities of acquisition devices along with the decreasing costs of processing and storage media have led to an explosion of visual information. Specialised as well as generic image and video databases have long been fundamental components of several application fields (e. g. astronomy, biomedical imaging, remote sensing, etc.) and fueled by technological advances, they continue to grow at an unprecedented rate. For instance, the annual production of an ordinary radiology department in an industrialised country hospital is estimated to be in the order of tens of terabytes [181], whereas the latest satellites are capable of producing the same amount of data in a matter of days.

Moreover, with the advent of affordable digital cameras and scanners, the general public has begun to contribute considerably to the ever increasing number and size of image collections, either online or offline, while numerous libraries have also initiated the digitisation of their visual material. As far as the world wide web (WWW) is concerned, the total number of images is estimated to be in the order of tens of billions (2006), accounting for an important part of the total online data volume.

Considering the growth rate of these inherently rich information depositories, robust and automated tools for their management, search and retrieval become essential for their effective exploitation. Similarly to their text-based counterparts, the contents of image and video collections need to be organised, stored, indexed and retrieved efficiently. The respective problem of textual data was solved with the help of sophisticated database management systems (DBMS). However, although text-based DBMS represent a long established domain, their theoretical principles are not easily adaptable to visual content as they were conceived and developed with only alphanumeric data in mind. This deficiency becomes most evident in the case of retrieval, since images due to their inherent visual nature, need to be accessed not only based on their description or any other accompanying information (e. g. file name), but on their content (e. g. visual features) as well.

The research for fast and effective methods for content-based retrieval of visual data continues since the early 90's and represents a very active field with new publications on a daily basis. In this chapter, in an attempt to summarise the numerous methodologies and techniques that have been developed as a result of these efforts, and continue to be refined, a short overview of the general field of content-based image retrieval (CBIR) is provided.

Although the term "visual content" includes also videos, the scope of this chapter is limited with still images; nevertheless, as it will be detailed further on, the problem of video retrieval can be resolved with techniques of image retrieval once motion is ignored. We start by introducing the fundamental concepts behind any CBIR system and further elaborate on its underlying components (Sec. 5.2). Then we review some of the popular commercial and research CBIR systems (Sect. 5.3).

5.2 Image retrieval

The general problem of image retrieval can be stated as retrieving images relevant to the user's queries from a large collection, with the form of the query as well as the notion of relevance depending on the application at hand. It dates back to the 1970's, and has been a very active research field ever since. It combines the disciplines of two major research communities, namely Database Management and Computer Vision, with each of them following its own direction to tackle with the common problem, respectively the text-based and visual-based approaches [232].

The initially popular text-based approach, consists in adapting the visual data to the existing text-based database systems. Specifically, images are first annotated by means of their textual descriptions, chosen with the help of experts, and then indexed according to these keywords, which are in their turn used by the underlying text-based DBMS with the end of retrieval. Various methods were proposed on this topic, for a comprehensive survey of which the reader can consult Ref. [36].

However, although the text-based approach relies on a simple and efficient principle, its efficiency is limited only with relatively small collections, due to two major drawbacks. First of all and most importantly, considering the general complexity of visual content, each person perceives a given image differently, and thus the association of the same image with differing descriptions is quite possible. This lack of precision takes a serious toll on the subsequent retrieval process, with the disruption related to the subjectivity of the expert becoming all the more acute as the collection size increases.

Additionally, as noticed by Picard in Ref. [208], if every picture was indeed worth just a thousand words, then the retrieval problem would be immediately resolved, since it would be sufficient to index every image based on this set of words. But the crucial question remains unanswered, "is there an unique set of words describing perfectly every image, and if yes, who will determine them for all the images in the world ?". Secondly, even if the answer to the previous question is affirmative and such an expert is available, given a moderate collection of just a few tens of thousands of images, the vast manual labor and time required for annotation renders this approach prohibitive for most applications.

5.2.1 Exploiting visual content

The difficulties related to text-based image retrieval combined with the rapidly increasing image collection sizes were the main reasons behind the proposal of content-based image retrieval during the early 90's. The term CBIR, having been first used by T. Kato in 1992 [130], denotes the process of image retrieval based on their visual content, instead of text-based keywords. In essence, images are described by multi-dimensional feature vectors computed using visual characteristics such as colour, shape and texture, while retrieval is performed based on the similarities/distances of these vectors.

Since then, CBIR has become rapidly a thriving field, by capturing the attention of both the industry and the research communities, thanks to its broad application spectrum. Among several special issues of journals that have been dedicated to this subject, a decade ago Gudivada and Raghavan [90] identified a list of twelve application areas, that can be updated to include art galleries and museum management, architectural and engineering design, interior design, remote sensing and management of earth resources, geographic information systems, scientific database management, weather forecasting, retailing, fabric and fashion design, trademark and copyright database management, law enforcement and criminal investigation, home photo collections, communication systems, education and training, WWW searching, medical diagnosis, journalism and advertising [159, 202, 300].

Initial efforts in CBIR, were mostly concerned with establishing the feasibility of retrieval from large visual databases, based on automatically obtained visual features. During this early period, which spans the early and mid 90's, efforts were oriented mainly towards the two basic components of any CBIR system: the description of the content of a given image or feature extraction, and the calculation of the similarity between images. As a result countless visual features, using mainly colour, texture and shape, as well as matching techniques were proposed, which in their turn led to several commercial and research CBIR systems (e. g. QBIC [74], Photobook [206], VisualSEEk [248], Virage [19], Four Eyes [177], etc). For comprehensive reviews of the papers representing this period the reader may consult Refs. [117, 167].

Towards the end of this period which was marked with promising results, the field started to mature as the limitations of the previous models became clear. In particular, since the feasibility was demonstrated [66], efforts were devoted now to the satisfaction of user's needs, and specifically to the bridging of the *semantic gap* (Sec. 5.2.7), in other words the gap between high semantics and low level features. With this end, additional techniques were integrated into the CBIR systems, such as relevance feedback. Consequently, the number of fields participating in the CBIR community multiplied, as it is now studied by computer vision, image and video analysis, pattern recognition, artificial intelligence, multimedia database organisation, statistics, human-computer interaction, information retrieval, distributed computing and data mining. For exhaustive reviews of this latter period, spanning the late 90's, see Refs. [232, 247].

Since the year 2000, the number of researchers involved in CBIR continues to increase, as is the number of contributions per year. However, although the majority of the efforts are still oriented towards the core problem of image understanding, with CBIR becoming a cross-field, the transition that began at the late 90's, from generic to more specialised approaches, is now fully under way. Several new research directions have appeared, each concentrating specifically to a particular aspect of the CBIR architecture, while domain specific solutions also appear to be gaining interest. A broad review of the latest period can be found in Refs. [54, 146].

5.2.2 CBIR system architecture

Despite the enormous variety of techniques and methodologies for CBIR, the general architecture of CBIR systems has changed little during the last years. A diagram of a generic CBIR system is shown in Fig. 5.1. According to Ref. [56], several components can be identified within this structure:

- A mechanism for the extraction of features (either low or high level) from visual data, that will subsequently be used for their characterisation.

- A visualisation tool, providing a panoramic view of the visual information space.
- A graphical query tool, providing the interface to specify the visual clues that are sought by the user.
- An index structure, filtering out the images that are irrelevant to the user's query, since sequential scanning is impractical with very large collections.
- A retrieval engine, which depending on the context of the application, employs some similarity measure(s) to match the images corresponding to the user's query.
- A relevance feedback subsystem, where the user can indicate the relevance of the retrieved items, thus refining progressively the search results.

In the sequel of this section, we will elaborate on the main underlying components of CBIR systems, starting with feature extraction.

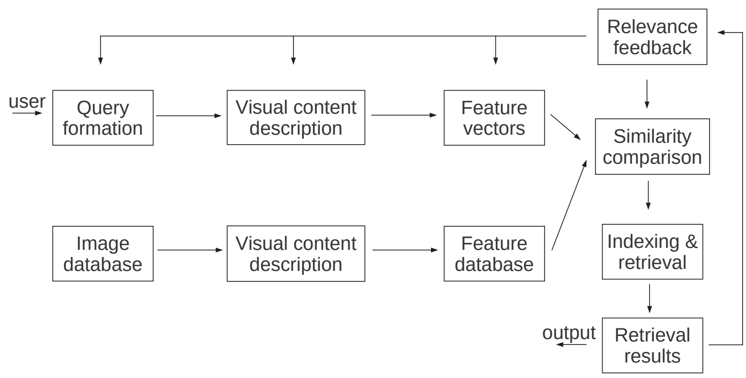


Figure 5.1: Diagram of a generic content-based image retrieval system [72].

5.2.3 Feature extraction

Feature extraction has received extensive attention since the very early days of image retrieval, since it constitutes a fundamental part of any retrieval system. In brief, this processing step consists in obtaining a pertinent description of the data at hand, with the choice of optimal features depending on the particular application.

In general, visual content can be modeled as an hierarchy of abstractions. At the lowest level, there is the digital representation of the visual media, and since pixelwise comparison of images is both computationally expensive and an inadequate similarity measure, a higher layer of abstraction is necessary. The next level includes the so-called *primitive* or *low level features* that are relatively easily extracted, such as edges, curves and colour regions, which when combined and interpreted as objects with attributes can lead to even higher abstraction levels. At the end of this transition from quantitative to qualitative features we find the *logical or high level features*, in the form of semantic concepts, involving several objects in different relations as well as abstract notions, such as "peace" [67].

As far as the evaluation of the various descriptors is concerned, besides the quantitative criteria relative to their retrieval performance, there is also a number of qualitative criteria. The following properties are sought in general from all visual descriptors, independently from the particular application:

- Effectiveness: the capacity to capture the nuances of the input image, with its semantic properties constituting the ultimate goal.
- Efficiency: concerns the computational complexity of the feature extraction scheme. Online retrieval systems for instance, cannot afford anything slower than close to real-time feature extraction, whereas computation time is of little significance to offline indexation systems. Moreover, although one expects a positive correlation between feature extraction complexity and effectiveness, there are cases where it does not hold, such as for the dominant color MPEG-7 descriptor [192].
- Size: the length of the extracted feature is another crucial factor, as it has a direct influence on the similarity computation process, and hence on retrieval time. That is why usually dimension reduction techniques becomes mandatory, even though decreasing the size of a descriptor can have a negative effect on its effectiveness.
- Robustness: all descriptors are required to be robust against noise, and in general to "small" variations of image content.
- Invariance to transformations: besides noise robustness, invariance to certain geometric transformations like translation and rotation are also among the desired properties for visual descriptors. Invariance to the viewing angle and illumination conditions can also be added to this list, while invariance to scale is of importance only to colour and shape.

At this point, one should also mention a key issue in feature extraction, concerning the automation level of this process, since manual annotation is not feasible with large collections. Specifically, there is an inherent trade-off between the automation level and the degree of domain independence. More precisely, systems with fully automatic featuring, make use of the available domain specific a priori knowledge, thus becoming restricted to a particular type of data. On the other hand, in order to obtain absolute domain independence, while not sacrificing the pertinence of the features, human intervention of some form to the feature extraction process becomes often necessary. As far as the quality of an image feature is concerned, a second trade-off exists between its discriminative power and invariance to image variations and transformations such as illumination changes, scale, rotation, etc. As a matter of fact, in general the more a feature is invariant to the numerous transformations the less effective it becomes it terms of discriminating ability. For an in depth study of this aspect of feature extraction see Refs. [30, 246].

The vast majority of the CBIR systems proposed so far, make extensive use of low level features calculated either automatically or semi-automatically, with progressively more of them trying to reach higher abstraction levels by either combining these features, or enriching them semantically. Whereas the use of purely high level features is still in its infancy [66]. The issue of this (semantic) gap between the extracted low level features and the higher level needs of the users is treated in Sec. 5.2.7. Moreover, low level visual features can be distinguished as general and domain specific, with the former, being traditionally organised into three groups: colour, texture and shape related features, computed either locally or globally; while the latter includes application dependent descriptions, such as fingerprints and eigenfaces. In the sequel, we will examine in more detail each of the general feature groups.

Colour

Undoubtedly, colour is the most extensively used visual feature for image retrieval. Being one of the principal properties of any object as far as humans are concerned, it also possesses additional attractive characteristics, in terms of image description, such as relative robustness against image corruption and invariance with respect to scale, view angle and distance. Furthermore, since colour is represented as a three dimensional value, it also holds a higher discriminative potential compared to greyscale pixel values.

However, an important problematic aspect of colour related features is their sensitivity to illumination variance, a common phenomenon in real world images. To explain, very different image descriptions may be obtained from the relatively same contents depending on the illumination conditions, particularly if the luminance component is not processed separately (e. g. in RGB) [63]. For this reason, a preprocessing step aiming to ease the effect of varying illumination, is often applied before colour feature extraction [73].

In order to fully exploit the potential of colour, the choice of an appropriate space is of paramount importance [172]. Although several colour spaces with varying properties are available, among those that are most frequently employed in the context of image retrieval we can list RGB, $L^*a^*b^*$, $L^*u^*v^*$ and HLS, along with its variants HSV and HSI. However, no colour space has yet been agreed upon as ideal for this task, since each of them has its own advantages. Two of the major factors among others influencing the choice of the colour space in this context are perceptual uniformity and intuitiveness in terms of the human vision system (HVS). The former, referring to the direct relation between measured and perceived colour similarity, is a particularly desirable characteristic, favouring the two aforementioned CIE spaces, whereas the latter concerns the approximation of the human perception of chromaticity in terms of hue and saturation (Chapter 2). Besides the variety of colour spaces, there is also a plethora of colour descriptors proposed in the literature to choose from. Some of the commonly used colour descriptors are the following:

Colour histogram: Being the most traditional colour content representation, it effectively describes the statistical distribution of colours within a given image by quantising the colour space [257]. Besides its low computational cost, it is also robust against translation and rotation about the view axis. Furthermore, thanks to its tolerance to small content variations, it constitutes an efficient shot change detection tool for colour video [144].

On the other hand, the colour histogram suffers from two major drawbacks. First, it does not take into account any of the available spatial information. Consequently, two spatially distinct images may end up having identical histograms. To remedy this problem, which constitutes a serious issue especially in large databases, several counter measures were proposed. One group of methods with several variations, consists in computing a local histogram for each sub-area of the image (*image partitioning*), chosen according to some predefined scheme, e. g. regular rectangular partitions or even for the regions resulting from a previous segmentation step. Even though the precision of this approach can be improved with more sub-areas, a trade-off is to be made between their number and the computational complexity of the overall process.

Another measure against the waste of spatial information is to enrich the histogram with spatial properties (*histogram refinement*), classic examples of which are the joint histogram, which computes the joint density of several pixel features along with their intensity [205] and the spatial chromatic histogram [49]; which for the pixels of each colour, besides their number, takes also into account their spatial barycentre as well as standard deviation from it.

Secondly, the usual 8-bit colour depth, combined with three channels, yield an excessive number of bins from a computational point of view. Although generally speaking, more histogram bins mean more discriminating power, in practice a sub-quantisation appears to be often necessary for the optimum result [56]. Various methods have been used in order to reduce the bin number and obtain compact colour histograms, including, but not limited to, clustering, PCA, DFT, DCT as well as wavelet based approaches.

As to the similarity assessment of two normalised histograms H and H' with n bins, several possibilities are available; for instance the L_1 and L_2 distances, defined respectively as :

$$D_1(H,H') = \sum_{i=1}^{n} |H_i - H'_i| \tag{5.2.1}$$

$$D_2(H,H') = \Big(\sum_{i=1}^{n} (H_i - H'_i)^2\Big)^{1/2} \tag{5.2.2}$$

with H_i denoting the i^{th} bin of H. Nevertheless, these relatively simple metrics are not preferred often in practice as they depend largely on the number of bins, conversely to the *histogram intersection* [257]:

$$D_3(H,H') = \sum_{i=1}^{n} \min(H_i, H'_i) \tag{5.2.3}$$

which takes into account only the colours present in the images. Extensions to this basic definition have also appeared, such as the *incremental intersection* [257]. Further histogram similarity measures include the weighted Euclidean distance [190] and average colour distance [93]. Despite its inconveniences, the colour histogram remains one of the most widely employed colour featuring tools with numerous implementational variations [92, 122, 143].

Colour coherence vectors: Colour coherence vectors (CCV), once more a histogram based method, were proposed by Pass et al. in Ref. [204] with the end of integrating spatial information into colour histograms. More precisely, each bin of the CCV is constituted of two values, the number of coherent (α_i) and incoherent (β_i) pixels of this colour. Coherent are considered only those pixels that belong to a "large" consistently coloured region, with the size being a user defined threshold. Consequently, the CCV of an image I quantised to n colours can be expressed as:

$$\mathrm{CCV}(I) = \langle(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)\rangle \tag{5.2.4}$$

while $\langle(\alpha_1 + \beta_1), \ldots, (\alpha_n + \beta_n)\rangle$ represents the bin sequence of the colour histogram of the image. Their similarity is calculated with the following distance [56]:

$$D(CCV, CCV') = \sum_{i=1}^{n} \left(|\alpha_i - \alpha'_i| + |\beta_i - \beta'_i|\right) \tag{5.2.5}$$

In brief, by separating the pixels belonging to large coherent regions from the scattered ones, CCV offer a higher degree of discriminating power with respect to the colour histogram, which is also asserted by its often superior performance.

Colour correlogram: The colour correlogram was first introduced by Huang et al. in Ref. [115], as an improvement over the classic histogram, targeting its lack of spatial information by exploiting the spatial correlation of colour pairs. Technically, it consists of a three dimensional table, where the first two dimensions represent the colour pairs and the last their spatial distance. Given an image I, with I_{c_i} representing the set of pixels of colour c_i the correlogram is defined as [115] :

$$\gamma^k_{c_i,c_j}(I) = \operatorname*{Prob}_{p_1 \in I_{c_i}, p_2 \in I} \left[p_2 \in I_{c_j} | \, |p_1 - p_2| = k \right] \tag{5.2.6}$$

with $|\cdot|$ being a norm. Consequently, for any pixel of colour c_i, $\gamma^k_{c_i,c_j}$ represents the probability that a pixel at distance k is of colour c_j. From a practical point of view the correlogram captures effectively any spatial information while it also benefits from the advantages of the histogram, being relatively tolerant to changes in viewing positions, camera zooms, etc. It has been reported to provide better results with respect to the CCV, but at a higher computational and storage cost. That is why it is often employed as a *colour autocorrelogram*, where only the relations between identical colours are studied [115] :

$$\alpha^k_c(I) = \gamma^k_{c,c}(I) \tag{5.2.7}$$

As to its similarity measure, Ref. [115] proposes a “relative” L_1 distance based measure:

$$|I - I'|_\gamma = \sum_{1 \le i,j \le m, 1 \le k \le d} \frac{|\gamma^k_{c_i,c_j}(I) - \gamma^k_{c_i,c_j}(I')|}{1 + \gamma^k_{c_i,c_j}(I) + \gamma^k_{c_i,c_j}(I')} \tag{5.2.8}$$

where m is the number of possible colours and d the number of different distances considered, for instance 4 in Ref. [194].

Colour structure descriptor: Given the variety and number of descriptors that have appeared even during the early years of CBIR, it became obvious that some type of standard would be soon required. This task was undertaken by the MPEG-7 committee, that proposed a set of descriptors for extracting colour, shape and texture related information [165]. Among the colour descriptors of the MPEG-7 standard, the *colour structure descriptor* (CSD) was later shown to outperform its counterparts [191]. The CSD is basically an histogram variant, which describes colour at a global level while also providing information on its local distribution. Specifically, it consists in using a structuring element, the size of which depends on the input image, and then forming a colour histogram, where each bin represents the number of structuring elements in the image containing one or more pixels of this colour. The resulting feature vectors are compared using the the L_1 norm.

Colour distribution entropy: This is a relatively recently proposed colour descriptor [256], that is based on the annular colour histogram presented by Rao et al. [219]. In particular, supposing that A_i is the set of pixels of colour bin $1 \le i \le n$, let C_i and r_i be respectively its centroid and radius, as defined in Ref. [219]. Then, with C_i the centre, one can draw N concentric circles, with radius jr_i/N for each $1 \le j \le N$. In which case the annular colour histogram is given by $(|A_{i1}|, \ldots, |A_{iN}|)$, where $|A_{ij}|$ denotes the number of pixels in the circle j of colour bin i (Fig. 5.2). Consequently one can now form the normalised annular colour histogram $P_i = (P_{i1}, \ldots, P_{iN})$, where $P_{ij} = |A_{ij}|/|A_i|$. The colour distribution entropy (CDE) index can then be computed as $(h_1, E_1, \ldots, h_n, E_n)$, where h_i is the i^{th} colour histogram bin and E_i denotes the distribution entropy:

$$E_i = E(P_i) = -\sum_{j=1}^{N} P_{ij} \log_2(P_{ij}) \tag{5.2.9}$$

Thus, the entropy is employed in order to measure the spatial dispersion of every colour. The associated distance measure (d_{CDE}) between the CDE index of two images f and g is:

$$d_{CDE}(f,g) = \sum_{i=1}^{n} \min\left(h_i^f, h_i^g\right) \times \frac{\min\left(E_i^f, E_i^g\right)}{\max\left(E_i^f, E_i^g\right)} \tag{5.2.10}$$

Of course, colour descriptors are by no means limited to these few operators. Several others have been employed, such as the Haar wavelet transform [119], dominant colour descriptor [58], colour layout descriptor [165], colour sets [248], colour histogram sets [50], etc.

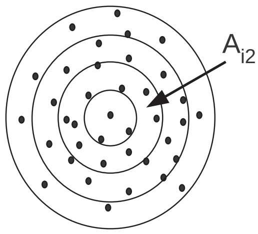


Figure 5.2: Concentric circles used in the construction of an annular color histogram.

Texture

Although the term texture has a relatively broad sense, defying any formal definition, it is widely accepted as a component of crucial importance for the human vision system, playing a vital role in surface and object recognition as well as in scene analysis. Generally speaking, texture refers to some basic visual pattern repeating itself in a semi-periodic manner. Given its importance and the large diversity of natural and artificial textures, it has been long investigated in the context of computer vision for more than three decades. There is practically an abundance of texture analysis and description methods, and most of them have already been applied to image retrieval.

Texture representation methods are usually categorised as *structural*, *model based*, *transform* and *statistical* approaches. Structural texture analysis (e. g. adjacency graphs, etc) consists in identifying the basic texture elements within a given image, and then exploiting their spatial placement rules with respect to each other. Structural measures concentrate on properties such as connectivity, density, etc. Thus, they are most adequate for describing highly regular textures. Model based methods on the other hand, attempt to interpret textures with particular models e. g. fractals, Markov random fields, etc. Transform techniques of texture analysis represent their input in a coordinate system where its pertinent features are more evident and easier to exploit. Popular tranforms of this type include Fourier, Gabor and wavelet transforms.

Statistical methods on the other hand, employ the intensity variation of the textures within a local window, and generally rely on second-order statistics computed using pixel pairs (e. g. co-occurrence matrices). Some of the atomic statistical texture features that have been proven to be effective in texture retrieval include contrast, directionality, regularity, coarseness [259], periodicity and randomness [153]. In the sequel, we elaborate on some of the texture representations frequently encountered in CBIR systems.

Tamura features: The *Tamura features*, including contrast, coarseness and directionality among others, represent basic properties of textures as far as the HVS is concerned. More precisely, contrast is related to the brightness distribution of the pixels and is defined as :

$$\frac{\sigma}{(\mu_4/\sigma^4)^{1/4}} \tag{5.2.11}$$

where σ is the standard deviation and μ_4 the fourth moment. Coarseness on the other hand denotes the texture's granularity. For its computation, first the moving averages $A(x, y)$ are calculated for every pixel over a window of size $2^k \times 2^k$:

$$A(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}} \sum_{j=y-2^{k-1}}^{y+2^{k-1}} p(i,j)/2^{2k} \tag{5.2.12}$$

with $p(i, j)$ being the pixel value at coordinates (i, j). Then, the k value maximising the differences (k_{max}) of the moving averages along horizontal and vertical directions is obtained for every pixel. Last, the coarseness results as the average of the k_{max} values over the entire image of size $m \times n$:

$$\frac{1}{m \times n} \sum_{i,j} k_{max}(i,j) \tag{5.2.13}$$

Instead of ending up with a single value for the entire image, improved versions of this feature result in a coarseness histogram, which can greatly boost retrieval efforts with textures possessing varying granularity patterns.

Furthermore, directionality measures how much a texture is concentrated along a certain direction. For its calculation, first the horizontal G_x and vertical G_y gradients are computed, within a 3×3 window. Hence, for each pixel the magnitude of the overall gradient $|G|$ and its direction θ are given by :

$$|G| = |G_x + G_y|/2 \tag{5.2.14}$$

$$\theta = \arctan(G_y/G_x) + \pi/2 \tag{5.2.15}$$

Subsequently, a direction histogram based on the values of θ, corresponding to "important" magnitudes of $|G|$, can be used in order to determine the dominant directions, if any, within the texture at hand. As far as their similarity assessment is concerned, Euclidean distances perform well with this type of descriptors.

Co-occurrence matrix: This technique exploits the spatial arrangement of pixels by means of mainly second order statistics. Although it is not suitable for textures composed of large patches [56], it is nevertheless one of the most widely employed texture descriptors. More precisely, for an image I and a displacement vector $\mathbf{d} = (dx, dy)^T$, each element $P_{\mathbf{d}}(i, j)$ of a co-occurrence matrix represents the occurrence frequency of a pair of pixels at grey values i and j, separated by $\mathbf{d}$:

$$P_{\mathbf{d}}(i,j) = |\{(q,r),(s,t) \in I \mid (s,t) = (q+dx, r+dy),\ I(q,r) = i \wedge I(s,t) = j\}| \,/\, |I| \tag{5.2.16}$$

where $|\cdot|$ denotes the cardinality of a set. Some of the statistic measures that are then computed from the resulting matrix are energy, contrast, entropy, homogeneity and correlation, defined respectively as :

$$\sum_{i,j} P_{\mathbf{d}}(i,j)^2 \tag{5.2.17}$$

$$\sum_{i,j}(i-j)^2 P_{\mathbf{d}}(i,j) \tag{5.2.18}$$

$$-\sum_{i,j} P_{\mathbf{d}}(i,j)\log_2 P_{\mathbf{d}}(i,j) \tag{5.2.19}$$

$$\sum_{i,j} \frac{P_{\mathbf{d}}(i,j)}{1+|i-j|} \tag{5.2.20}$$

$$\frac{\sum_{i,j}(i-\mu_x)(j-\mu_y)P_{\mathbf{d}}(i,j)}{\sigma_x\sigma_y} \tag{5.2.21}$$

where μ_x, μ_y are the means and σ_x, σ_y are the standard deviations of respectively $\sum_j P_{\mathbf{d}}(x,j)$ and $\sum_i P_{\mathbf{d}}(i,y)$. The main drawback of co-occurrence matrices is the lack of a reliable method for the choice of the displacement vector. Moreover, given the computational cost of matrix construction, the use of multiple vectors with this purpose is often disadvantageous.

Gabor filter features: Gabor filters (or wavelets) have been used for a variety of feature extraction tasks in image analysis, and particularly for texture description and segmentation [55]. They operate in the frequency domain, and their featuring capability relies on the distinction of low from high frequencies, representing respectively coarse and fine textures. Besides, they often exhibit superior retrieval performance compared to other multi-resolution features [164]. Several implementational variants of Gabor filters are available, as far as texture characterisation is concerned. Specifically, a Gabor function $g(x,y)$ is defined as:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right] \tag{5.2.22}$$

in other words, it consists of a sinusoidal plane wave modulated by a Gaussian envelope of frequency W, and σ_x with σ_y represent respectively the horizontal and vertical standard deviations of this envelope. Next, the self similar Gabor filters can be obtained through :

$$\begin{aligned} g_{mn}(x,y) &= a^{-2m}g(x',y') && (5.2.23)\\ x' &= a^{-2m}(x\cos\theta + y\sin\theta) && (5.2.24)\\ y' &= a^{-2m}(-x\sin\theta + y\cos\theta) && (5.2.25) \end{aligned}$$

where $m \in \{0,\ldots,M-1\}$, $n \in \{0,\ldots,N-1\}$, $a > 1$ and $\theta = n\pi/N$. N and M denote respectively the number of orientations and scales. Hence, given an image I, its discrete Gabor transform $G_{mn}(x,y)$ is given by :

$$G_{mn}(x,y) = \sum_{i,j} I(x-i,y-j)g^*_{mn}(i,j) \tag{5.2.26}$$

where $*$ indicates the complex conjugate. In practice the mean μ_{mn} and standard deviation σ_{mn} of the transform's magnitude are employed for texture description.

Variogram: From a morphological point of view, granulometry and covariance, detailed in Sec. 6.3, represent the main texture description tools. The variogram, on the other hand is a generalisation of morphological covariance that can be easily extended to colour images [101]. Specifically, in order to calculate the variogram of an image $f(\mathbf{x})$, a direction α and a unit displacement vector $\mathbf{h}$ in this direction must be chosen. For various multiples of $\mathbf{h}$, written $q\mathbf{h}$, the following value

$$V(q, \alpha) = \frac{1}{2}\mathcal{E}\left[f(\mathbf{x}) - f_\alpha(\mathbf{x} + q\mathbf{h})\right] \tag{5.2.27}$$

is plotted against q, where $f_\alpha(\mathbf{x} + q\mathbf{h})$ is the displacement of f in direction α by distance q, and $\mathcal{E}$ the expectation value. According to Ref. [101], its colour image implementation may be achieved by replacing the grey-level difference of Eq. 5.2.27 by the Euclidean distance in L*a*b*. Hence the variogram provides information on the directionality and periodicity of its input.

The aforementioned texture descriptor categories are not mutually exclusive, since mixed features unifying structural and statistical methods have also appeared, such as the local binary patterns [193]. Additional efforts in texture description concentrate on the exploitation of colour information for more effective results [162].

Shape

Given the higher degree of sensitivity of the HVS to luminance changes, compared to chromatic changes, edges and consequently shape is the primary characteristic employed by humans with the end of object identification. Although texture features also aim at exploiting statistical edge properties, shape features are more oriented towards deriving semantic edge properties, or in other words object boundaries [67]. However, shape based retrieval continues to be one of the hardest problems of general image retrieval. The inherent difficulty is due mainly to the initial need for detecting the objects' boundaries and the lack of a robust and automatic segmentation solution. Thus, shape based features perform poorly in the general case [277].

Besides, contrary to colour and texture, that are considered to be intensity based properties, shape is of geometric nature. As images only contain 2D projections of 3D objects, the problem of accurate description of shape is far from resolved. Furthermore, similarly to textures, invariance to geometric transformations is also essential. Faced with this challenge, extensive efforts have been made in the past, in order to effectively characterise the shape information and countless techniques have been developed. They are generally classified into two categories, *boundary based*, using only the edges of the objects, and *region based*, exploiting the entire object. The former group of methods is represented primarily by Fourier descriptors whereas the latter by moment invariants. These groups are further subdivided to organize the large variety of descriptors, including elementary geometric features (e. g. centroid, area, perimeter, orientation, eccentricity, circularity, curvature, etc), as well as more sophisticated approaches (e. g. template matching, Hough transform, etc). Further details on some of the widely used shape descriptors follow.

Moment invariants: Moments constitute well known pattern recognition tools, employed especially in shape analysis among other areas, despite their relatively high computational cost. They are most often used when the complete object of interest has been identified within an image. More precisely, the unscaled moment m_{pq} of order (p, q) of a greyscale image

f of size $M \times N$ pixels is given by :

$$m_{pq}(f) = \sum_{x=1}^{M} \sum_{y=1}^{N} x^p y^q f(x,y) \tag{5.2.28}$$

Yet moments are seldom employed in this form; in practice the *moment invariants* are preferred [114], that are simply defined as functions based on equation (5.2.28), and exhibit some kind of invariance with respect to certain transformations, such as translation, rotation, scale, etc. Most usually only the first few orders are computed, since higher order moments tend to be relatively noisy.

Fourier descriptors: In contrast to moments invariants, the coefficients of the 2D discrete Fourier transform are employed for the description of object boundaries. Specifically, the 2D discrete Fourier transform of an $M \times N$ image f is defined as :

$$F(m,n) = \sum_{x,y} f(x,y) e^{-2\pi j(mx/M + ny/N)} \tag{5.2.29}$$

Consequently, by passing to the frequency domain, one can consider only the first few coefficients that account for the lower frequencies and represent the "important" edges. Hence a series of translation invariant shape features is obtained. As to their similarity assessment, several alternatives are available including the standard Euclidean distance as well as more fine tuned choices such as Dynamic Time Warping that exploits both the amplitude and the phase [22].The coefficients of the 2D discrete wavelet transform can also be used in a similar manner, but with the additional advantage of multiscale representation.

The discriminative potential of shape based features however is seriously limited by the performance of the segmentation method at hand, rendering them useful particularly for images of low complexity with well distinguished objects, or application specific cases where external knowledge facilitating the segmentation step is available.

Local invariants

Local invariants, like for instance interest or corners points, that are considered important tools in stereo vision for matching objects taken at various views, have gained a considerable momentum in the last few years, as far as feature extraction for CBIR is concerned. In principle, by limiting the feature extraction process to the immediate neighborhood of a set of *representative points*, they provide a means for describing the visual content compactly and robustly against view and illumination changes. To explain, traditional interest point detectors, such as the Harris detector [107], compute the corners of the objects that represent salient and representative points of the object under consideration. Hence a characterisation of for instance the edge orientations in the neighborhood of each point can offer a pertinent description of the object. A considerable problem however is scale invariance, since the Harris detector is sensitive to scale changes. To this end, automatic scale selection algorithms have been developed, thus leading to descriptors of formidable performance, such as the Harris-Laplace detector [176], where the Laplace operator is used to select the scale at which the Harris operator extracts the interest points. A survey on the different types of local invariants as well as a comparative study may be found in Ref. [238].

Spatial information

Although not as popular as the preceding feature types, spatial information represents an important visual characteristic of image content that is often employed in combination with the previous three. The approaches to the exploitation of the inherent spatial information of images with the end of retrieval, concentrate mainly either on the spatial location of image components (objects or regions), or on their spatial relationships (e. g. X on top of Y, X overlap Y, etc). However, in order to reason in terms of objects and regions, a symbolic image denoting those components is necessary. That is why, similarly to shape based features, the use of spatial information in this manner also requires a reliable segmentation. Consequently, spatial features have appeared most often in application specific retrieval systems where robust segmentation is feasible, for instance in medical imaging, where a wealth of a priori knowledge is available [46]. Additionally, as a way of circumventing the need for a segmentation, alternative approaches were proposed involving the division of images into sub-areas according to a predefined scheme. Yet their efficiency is debatable since the spatial organisation of real world images is too diverse to be fully captured by a single model.

Non visual features

Non visual features are most usually of textual nature. There are various application areas where text based information can be at least as important as the visual content of the images. For instance in photographic collections, accompanying information such as date, occasion and acquisition conditions are of critical importance. Besides their use as auxiliary meta-data, textual descriptions are also employed in the form of captions and annotations. Text based description of visual information dates of course to the early years of CBIR, but despite its inconveniences (Sec. 5.2), it remains popular. Primarily because semantic queries are easier to formulate with keywords and due to the fact that end users are already accustomed to text based retrieval with popular search engines such as Google and Yahoo!. Considerable effort was made in the early years in order to remedy the disadvantages of text-based descriptions, and particularly the subjectivity of the expert. Approaches such as creating a visual thesaurus and establishing specific description templates have appeared [16, 208].

Within the new age of CBIR, two approaches have distinguished themselves in this direction. In the first case we have *region based textual annotation*, which is briefly based on the association of image regions with keywords. More precisely, this *translation model* is adopted by Duygulu et al. [65], where images are first segmented into well defined regions called *blobs*, and then each region is described in terms of visual features. The challenge lies in the association at this step of each region signature with a corresponding keyword, chosen from a predefined vocabulary, based on statistical methods and previous learning sessions. In other words a translation takes place, at the end of which every image is described by a set of words. Extensions to this model appeared in Refs. [26, 124] that refined the association process. Cross-media relevance models [121, 140] have further improved this approach's performance by exploiting word and blob correlations [123]. Variations based on probabilistic latent analysis [154, 180] as well as multiple instance learning [296] have appeared. Although promising results have been obtained with blob-word associations, this approach fails to incorporate any spatial relations among the blobs, while its main drawback is its dependence on a reliable segmentation.

On the other hand, *supervised classification* based methods, treat every image as a whole, while attempting to associate it with a text label. In particular, mainly statistical approaches are employed in order to model the relations between labels and global visual features. After the learning process, it becomes a mere likelihood computation to determine which label is "closest" to the image at hand. Various techniques have been employed within this context, including hidden Markov models [151], Gaussian mixture models [60], Bayes point machines [34], SVM classifiers [148] as well as multiple instance learning [42]. Although both approaches make the assumption that there exists a set of words for describing an image, the second alternative is less flexible as it associates the entire image with a single keyword, however it does not depend on a segmentation step. Furthermore, since purely content-based solutions are still far from ideal in the general case, textual captions and annotations continue to be invaluable assets for image retrieval.

In brief, a plethora of features are available, however as no fixed set of features can be used to describe effectively all images, the choice of an appropriate set still remains ad-hoc. AI based methods of feature choice appear to be gaining interest in this regard [66]. Moreover, considering that important ground has been covered during the last decade in low level feature extraction schemes, efforts have now been relatively shifted towards their exploitation and effective combination, with the end of satisfying the high level user needs and bridging the semantic gap.

5.2.4 Query formulation

Retrieval sessions usually begin with the formulation of a query, describing the sought content. Considering the potential complexity of images as well as the diversity of user needs, several query types have been reported in the literature. Initially they were classified as queries based on low level (e. g. colour, texture, shape) and high level features (object or event illustrated) [90]. The distinction was later further refined [66], by identifying three basic levels of queries, with each of them targeting a particular abstraction level of image features.

According to this classification, the first level queries make direct use of primitive features. For instance, "retrieve images containing 35% of green", or most likely combinations of such quantitative features, involving numerous properties, e. g. "retrieve images with blue on top, X texture in middle, round objects in bottom and acquired on a Wednesday". In the second level, we find queries of objects and concepts, such as "cathedral" as well as groups of them in spatial or in other types of relationship, e. g. "retrieve images of children entering school". Whereas the third level corresponds to abstract notions involving complex reasoning, for example "retrieve images of war". The vast majority of the so far presented retrieval systems offer exclusively first level queries. However, user needs have long been identified mainly on the second and third levels [15]. Although experimentation efforts continue on second level queries, the complexity of the third renders it out of reach for the time being.

Naturally, from an implementational point of view, the exact query format depends on the application. To illustrate this concept, the aforementioned query types can be formulated in a text based manner directly by the user or by means of predefined values specified by the graphical user interface. Nevertheless, since the average user does not reason in terms of exact colour proportions nor is an image processing expert, visual queries have become popular in this regard. In particular, they allow the user to formulate a visual query, for instance by sketching the desired shape (query by sketch), by 3D models or by providing indications on its histogram (query by histogram). Another form is to query with the help of an example image (query by example), with or without a particular region of interest, in which case it is totally up to the feature extraction system to decide what description to associate with it.

The last case appears to be particularly popular as it permits the formulation of conceptual queries, for instance by providing the image of a sunset, without the need for subjective descriptions, that may vary from user to user, e. g. “a sunset”, “view of the sun going down”, “the sky in the afternoon”, etc. On the other hand, although the user may have asked for a sunset, the perception of the retrieval system is still limited by its own recognition capabilities, and there is no guarantee whatsoever that its perception of the content will match that of the user's. Hence, depending on its sophistication, the system can easily return images containing just large amounts of red and yellow, that may or may not illustrate actual sunsets.

5.2.5 Indexing and similarity assessment

Effective similarity assessment is of vital importance to any CBIR system. Even if the chosen features are of superior quality, the overall performance of the system is still limited by its capacity to effectively compare them. Additionally, considering the current sizes of visual databases, as well as the multitude of feature vector dimensions, usually in the order of hundreds, sophisticated indexing systems for fast access and retrieval of data are an absolute necessity. Consequently multidimensional indexing and similarity measuring approaches have formed a separate research direction in the CBIR field, with numerous variants.

With even common databases containing thousands of items, sequential scanning is to be avoided with the exception of relatively small collections. For the rest, an efficient indexing scheme is a must. An organisation of such a kind is needed that will effectively filter out all irrelevant images and rapidly locate the group most relevant to the user's query. Several data structures have been employed with this end during the last decade, including quad-tree, K-D-B tree, and especially R-tree and its variations. Lately, AI related methodologies are also being extensively investigated in this context, e. g. neural nets, clustering algorithms, self organising maps, etc. Moreover, as image databases are dynamic structures, an additional important requirement on the indexing infrastructure is the ability to add new images without having to update the entire system. As far as the high dimension of feature vectors is concerned, a way of simplification is to apply a dimension reduction technique using any of the available transforms, such as the widely employed PCA, with variants facilitating the addition of new images and column wise clustering, among others. Of course, the effect of the lesser amount of features on the end results is unavoidable, but according to Ref. [290], their pertinence is only slightly degraded.

Given a set of effective features with reduced dimensions, along with a robust indexing system, another necessary component before attempting retrieval is a means of computing image similarity. Considering the vector nature of features, practically any distance measure can be employed with this end. However, the ideal solution to this problem is far from trivial, as the similarity assessment of the HVS is quite complex, hence numerous similarity models have been tested so far with varying performances in attempts to approximate it [56].

5.2.6 Video indexation

Similarly to still images, with the domination of visual information, large video databases have also been formed, especially in domains related to visual entertainment, tv broadcasting, journalism, surveillance systems, etc. While humans are capable of almost instant content perception of image data, much more time and effort is required for videos, hence the need for their effective retrieval is even more acute. Furthermore, by forming a combination of spatial and temporal information, videos demand great amounts of memory for their storage and processing, and that is why during the initial years of CBIR they did not receive much attention. But the situation rapidly changed, thanks to the advent of high end and cost effective computing systems, that have contributed to the increasing popularity of the area of video retrieval.

The difference of videos with respect to images is the temporal dimension and optionally the presence of audio information. Although sound holds important discriminative potential we will limit our scope with the visual aspect of the issue. The availability of multiple frames, is exploited in the context of content based retrieval as an aid for object recognition as well as for the formulation of motion based queries [129]. Apart from that, the efforts made in video retrieval have been mainly concentrated in adapting the existing image retrieval techniques, by first reducing the video data to a smaller set of still images.

Video is considered as a set of segments, i. e. *shots*, *scenes* and *episodes*. A shot is a sequentially recorded series of frames by a single camera containing some continuous action, whereas a sequence of shots focusing on the same event or location is called a scene. An episode on the other hand is a set of related scenes [288]. Thus, by extracting representative frames from each shot, i. e. *keyframes*, one can obtain a set of still images summarising the visual content of the initial video, upon which the existing CBIR methods can be applied. Of course, the effectiveness of this approach depends critically on the successful structural segmentation of video into its constituent segments, which is further complicated with the presence of transition and other types of commonly used effects. Numerous techniques have been proposed with this end, making extensive use of histogram comparisons, for a concise survey on this topic see Ref. [144].

5.2.7 Semantic gap

To the best of the author's knowledge, the term "semantic gap" was first used in this context by Del Bimbo in his book, "Visual Information Retrieval" (1999) [56]. In a broader sense, it refers to the mismatch that appears in layered systems, when high level abstract concepts need to be translated into lower concrete artifacts. For instance in computer architecture, it is used to denote the difference between the complex operations performed by high level constructs and the low level instructions. Likewise, in CBIR, it describes the difference between the previously presented low level visual features and high level notions used by humans.

Although several CBIR systems have been developed, equipped with efficient low level featuring and querying systems, their appeal to the end user is still limited, since their needs lie in the upper two query levels. This gap was quickly noticed and beginning from the mid-90's, efforts to bridge it are being made. Since no information is yet available on the human behaviour as far as the recognition of third level abstract notions is concerned, all efforts in this regard have been almost exclusively oriented towards second level semantic retrieval; the core problem of which consists of object recognition and context identification. According to Ref. [66], the approaches in this direction can be classified into the following categories.

Automatic scene classification

This type of classification into general groups such as indoors, outdoors, city scenes and landscapes, can be highly rewarding as a preliminary step in filtering out large irrelevant portions of a database with respect to a given query. Typically statistical techniques have been adopted for this approach, coupled with various low level features for characterising the image content [270]. In the case of general image collections, unsupervised clustering methods are usually employed with the end of organising the database content into "meaningful" clusters, hence accelerating the retrieval process. Although they require only minimal user intervention, the clustering result strongly depends on the same low level features. Specifically, some of the representative papers on this topic include Refs. [43, 302]. Further WWW oriented application examples benefiting from the availability of text based meta-data can be found in Refs. [79, 285].

More sophisticated classification techniques become advantageous in the case of domain specific collections, where the nature of labels is known beforehand, and simple training sessions can be considerably beneficial. Among the employed approaches we can cite Bayesian classifiers, k-NN and SVM [42]. On the other hand, one can never be sure of having captured all the images classes, since training images unavoidably introduce some degree of bias. Although scene classification represents certainly a step towards semantic retrieval, the so far obtained results are still far from ground-breaking, as the employed number of labels is generally severely limited; besides their validity is hard to evaluate as it strongly depends on the image database at hand.

Automatic object recognition

Object recognition is of paramount importance for semantic indexing, as it denotes an elementary step in image understanding. The two primary methodologies employed with the end of automatic object recognition are distinguished as knowledge based and statistical. The former approach is based upon the use of models describing the desired object, which is later sought for within the image. For classic applications of this technique see Refs. [64, 76]. Thanks to the use of AI techniques and heuristics, knowledge based approaches exhibit a certain level of reasoning, the efficiency of which however is strictly limited with the application specific, high level object model provided by an expert. Yet, although the diversity of objects in real world images is daunting, object models certainly represent a promising research direction.

Statistical methods on the other hand, likewise scene classification techniques, do not rely on any predefined object models. Instead, they exploit the statistical relations, if any, between low level visual features and object semantics, obtained by means of training sessions. Examples of this approach may be found in Refs. [168, 271]. Despite the absence of any high level notions, statistical techniques do have the advantage of not needing any domain dependent object models.

User assisted retrieval techniques

The aforementioned two approaches, although fully automatic, are still far from capturing the subtleties of human visual perception. Besides they incorporate no learning mechanisms, except for an eventual initial training, whereas the notion of adaptability is crucial for high level vision and the satisfaction of varying user needs. The concept of learning can be integrated into image retrieval either by statistical means, which is the case with object and scene recognition, or with human intervention [231]. The latter option has been investigated extensively during the last decade and various methodologies were developed with this purpose.

The majority of user assisted retrieval techniques is based on the principle of *relevance feedback*, a notion belonging to the field of information retrieval, and introduced into CBIR in Ref. [234]. More precisely, human intervention occurs with the judgement of the user, whether the retrieved images are relevant or not and to what degree, to her query. Thus, thanks to the instant visual content analysis capacity of the HVS, semantic information can be rapidly obtained in order to improve the future retrievals. Since there is no general purpose effective semantic framework, relevance feedback constitutes a relatively easy way of obtaining case specific semantics. Several types of feedback mechanisms have appeared in the CBIR literature, for a concise survey on this topic see Ref. [303].

Feedback based retrieval systems appear to vary primarily on the type of information provided by the user, its exploitation within the system and on the learning process implementation. For instance, while some systems support only binary input from the user, either in the form of only positive examples [44] or accompanied by negatives too [262], others accept even fuzzy feedback [294] and multiple groups [113]. Furthermore, as far as region based systems are concerned [33], since some images may contain simultaneously relevant and irrelevant parts, region based feedback has also been studied [125]. Instead of images, users can additionally express their opinions in terms of labels, or even in combination with textual data [297].

The information returned by the user can be employed in a variety of ways depending on the system's architecture, e. g. in order to update the similarity model [77], the association of features to query templates [35], etc. The main concern however is to establish the relation between low level features and high level semantic information, relying exclusively on the user. In practice the learning process can be realised with various techniques, including feature and region weighting, active learning, boosting, etc [82, 108, 152], whereas probabilistic methods have also been used with this end [53, 276]. In addition, certain optimisation techniques have been proposed in this regard, rendering relevance feedback an even more effective tool [94].

CBIR systems employing user assisted retrieval techniques are often cited to have superior performances, even though comparative evaluation attempts are faced with serious difficulties in this field (Sec. 5.3.1). Assuming that the users are capable of identifying the semantic content of images, in this case the question of effective retrieval becomes equivalent to properly exploiting this feedback in order to improve the system's capabilities. Diverse AI based methods are progressively examined in this regard [41], with various aspects still untouched [275]. Moreover, user assisted systems can continue their learning and self improvement indefinitely, nevertheless, the human factor is also their weak point. Since they do not benefit from any higher level knowledge that is built into model based systems, they are totally dependent on the quality of their interaction with the users.

5.3 CBIR Systems

By now it should be clear that image retrieval is a very active research domain with new developments occurring on a daily basis. A long way has been covered since the early years, and several commercial and research systems have been developed, each with distinct characteristics and capabilities. Moreover, the objectives of researchers have also evolved, starting with mere feasibility tests and advancing through various stages, getting progressively more oriented towards user needs. The users' needs as well as the nature of the image collections, continue being today the primary factors influencing research directions. This abundance however of different architectures and techniques, has also brought up the issue of evaluation. How to determine which one is the best, or at least better, compared to some others? In this section, we will first elaborate on the issue of performance evaluation, and then review the main requirements of real world CBIR systems, followed by a brief introduction of certain classic CBIR systems; for a comprehensive review on this topic the reader can consult Ref. [278].

5.3.1 Performance evaluation

The objective evaluation of CBIR systems' performance continues to pose problems even today. The main issues concern the difference of test images, the variation of performance measures as well as the overall homogeneity and nature of the image collection at hand. Similarly to the domain of information retrieval, the traditional measures of performance in CBIR are *recall* and *precision*. The former denotes the percentage of relevant images retrieved from the database, whereas the latter represents the percentage of images relevant to the query among the retrieved ones. Specifically they are defined as :

$$\text{Recall} = \frac{\text{number of relevant and retrieved images}}{\text{number of relevant images}} \tag{5.3.1}$$

and

$$\text{Precision} = \frac{\text{number of relevant and retrieved images}}{\text{number of retrieved images}} \tag{5.3.2}$$

Of course both measures require a ground truth, in other words it must be a priori known which images are relevant to the employed query. In an attempt to produce an objective performance measure, the MPEG-7 standard has proposed the *average normalised modified retrieval rank* (ANMRR), which has been accepted relatively widely by the CBIR community [165]. More precisely, this measure has been introduced in order to overcome problems such as those related to queries with varying ground truth set sizes. In particular, given a query q, assume that the k^{th} ground truth image is retrieved at $\text{Rank}(k)$. Then a number $K(q)$ is defined, which denotes the ranks that are considered as feasible in terms of retrieval evaluation. $K(q)$ is often set as twice the size of the ground truth set (NG) associated with the query under consideration. Hence $\text{Rank}^*(k)$ is defined, which represents the penalty attributed to a retrieved item:

$$\text{Rank}^*(k) = \begin{cases} \text{Rank}(k), & \text{if } \text{Rank}(k) \leq K(q) \\ 1.25K(q), & \text{if } \text{Rank}(k) > K(q) \end{cases} \tag{5.3.3}$$

Thus one gets the average rank (AVR) for a query q:

$$\text{AVR}(q) = \frac{1}{\text{NG}(q)} \sum_{k=1}^{\text{NG}(q)} \text{Rank}^*(k) \tag{5.3.4}$$

And in order to minimise the effect of different ground truth sizes, the modified retrieval rank (MRR) is defined:

$$\text{MRR}(q) = \text{AVR}(q) - 0.5(1 + \text{NG}(q)) \tag{5.3.5}$$

which is then normalised to give the normalised modified retrieval rank (NMRR):

$$\mathrm{NMRR}(q) = \frac{\mathrm{MRR}(q)}{1.25K - 0.5(1 + \mathrm{NG}(q))} \tag{5.3.6}$$

Thus NMRR can provide values between 0 (i. e. all images of the ground truth of q have been retrieved) and 1 (i. e. none of the ground truth images of q has been retrieved). Consequenly, one reaches the definition of ANMRR simply by averaging the NMRR values over all queries:

$$\mathrm{ANMRR} = \frac{1}{\mathrm{NQ}} \sum_{q=1}^{NQ} \mathrm{NMRR}(q) \tag{5.3.7}$$

where NQ is the number of queries.

Initially Corel Photos were quite popular as a common evaluation collection, however unintentional configuration choices of classes have been shown to lead to misleading results [183], and hence new datasets have emerged, such as Coil [189], PASCAL VOC [69], Caltech-101 & -256 [149, 88] and Graz [195]. Moreover, evaluation campaigns such as ImageCLEF [118] have also met with acceptance. Considerable efforts have been recently made in the video indexing field with the organisation of annual benchmarking exercises such as TRECVID [245]. A survey of the different evaluation techniques and algorithms can be found in Ref. [184].

5.3.2 Real world requirements

Given the variety of applications domains where image retrieval has been adopted, usually as a fundamental tool, different architecture and organisation configurations become often necessary in order to obtain the optimum result. The following list of critical issues, as far as real world deployment of CBIR systems is concerned, was compiled in Ref. [54].

- The performance of an image retrieval system is naturally of paramount importance. So far, the results obtained for general image collections have been relatively poor, compared to application specific cases. Besides, if the retrieval quality is not sufficient, neither commercialisation nor extended public use can be expected, hence major efforts are being made for boosting the performance of the current architectures.

- Semantic retrieval represents the current major bottleneck of CBIR systems, which has led researchers to orient themselves particularly towards alternative and particularly AI related methodologies. Learning is regarded as necessary in this end, with relevance feedback having become a standard component of the CBIR architecture. The daunting complexity of semantic retrieval has lately favoured the development of domain specific systems, where the availability of external knowledge and low visual content diversity facilitates the partial resolution of the problem.

- The variety of data to retrieve, this time not in terms of content, but rather in terms of acquisition source and technical properties, e. g. number of colours, resolution, etc, constitutes a further unavoidable design challenge to be undertaken in order to create a general purpose CBIR system. Yet, the temporal dimension of videos and the information carried by sound, are also two additional components to process, since videos nowadays are as common as images in most visual databases.

- A system's response time is a crucial property as far as the end user is concerned, especially considering their familiarity with the almost instant text based search engines. Thus, operating speed becomes a significant parameter, and distributed architectures permitting the concurrent use of the system necessary. Besides that is why feature extraction is most usually realised offline, in other words, before putting the system to user disposal, conversely to online annotation, where the features are computed after the query formulation. For further speed optimisation, features computed in a compressed domain are favoured, while hardware support is also among the retrieval acceleration options [137].

- Besides the speed of retrieval, CBIR systems need additionally to take into account the potentially huge volume of data, requiring robust designs and effective indexing organisations. This need is even more acute for video databases, where the data volume is usually much higher.

- Given the yet limited use of CBIR systems by the general public, user interface design carries considerable importance, since it is the main interaction interface with the end user. For instance even manual annotation becomes attractive if realised through a game [283]. Intuitiveness and ease of use are critical notions in this case, especially as far as query formulation is concerned, since the vast majority is not prone to define colour percentages. Visualisation of the results is also part of the same problem and based on how people manage their own collections, various approaches have been tested, including spiral and concentric formations [295].

- The comparative evaluation of CBIR systems and of the different competing approaches is still problematic, for the reasons given in Sec. 5.3.1. An objective benchmark and common measures are necessary in order to set some industry standards. The ISO standard MPEG-7 [165] as well as benchmarking exercises such as TRECVID [245] represent positive steps in this direction.

5.3.3 Commercial systems

The number of commercially available CBIR systems is relatively low compared to research systems. As a matter of fact their number asserts their lack of appeal to the general public, as their capabilities are currently far from satisfying the mainly semantic needs of users. Nevertheless, even in this state they can be of great use in several areas. Some of the most widely known commercial systems include QBIC and Marvel by IBM Corporation, VIR Image Engine by Virage Inc. and Excalibur Visual RetrievalWare from Excalibur Technologies.

Query By Image Content (QBIC) [74] is a rich framework incorporating several features and options, with additional support for video content. Colour is represented by means of 3D average vectors in 4 spaces (RGB, YIQ, Munsell, Lab) and a 256-dimensional RGB histogram. Moreover, modified versions of the Tamura features are employed for texture characterisation. Shape and spatial layout similarity is also supported. Besides the basic descriptors it also supports semantic querying. As to the videos, automatic segmentation is also among the available features. The indexing of the feature vectors is realized with the help of R^*-trees and last, sketch and text-based queries are also possible besides the specification of standard visual descriptors. Its application to the search of world famous art archives can be found in Ref. [216].

The successor of QBIC within IBM corporation, as their leading research project in this field, has been Marvel [169]. Based on the MPEG-7 standard, Marvel concentrates on the automatic extraction of up to 200 higher level semantic concepts. However, it only takes into account standalone concepts while inter-concept relations are ignored.

The VIR Image Engine [91], likewise QBIC supports the querying of both still images and video content, however it was designed for building CBIR systems. It possesses an extensible structure with four abstract feature types: global and local values and histograms along with graphs. Hence, besides the feature set that is provided with the system, a developer can easily add her own. It supports querying by semantic content, by global colour similarity, by texture and by structure similarity, as well as their combinations. Sound based video retrieval is also supported along with other standard temporal functions. The VIR system makes use of weight based relevance feedback whereas the indexing and storage of the feature sets is the developer's responsibility.

Visual Retrievalware [62] provides a framework for the retrieval of only still images. Making use of neural nets, it supports semantic querying and query by colour and texture similarity. It also permits the weighted combination of these primitive features by means of an intuitive GUI.

5.3.4 Research systems

The following list includes some of the CBIR systems that were developed in universities and research laboratories.

The Photobook system, developed at MIT Media Lab [206], is constituted of three sub-systems, each oriented towards the description of a specific type of images, namely faces, shapes and textures. Hence, depending on the query at hand, the image specific features can be effectively employed for retrieval. The FourEyes system, a later version of Photobook, includes also a learning mechanism based on human intervention during the annotation and retrieval stages.

The VisualSEEk system [248], along with its WWW oriented version WebSEEk [249], were both developed at Columbia University. Their main characteristics include spatial relationship based image region description and query. Given that most online images are available in a compressed form as well as their enormous number, WebSEEk makes use of compressed domain visual features along with an optimised binary tree based indexing organisation.

The NeTra system was developed at the University of California [160]. It has been designed to support queries by colour, texture, shape and spatial location with features computed based on image regions. In particular, it includes Gabor filter based texture analysis and neural net based image thesaurus construction.

The MARS (Multimedia Analysis and Retrieval System) framework is developed at the University of Illinois [116, 233], and what makes it different from most of its counterparts is that it acknowledges the absence of one unique set of features best describing every image, and hence attempts to organise the various visual features in such a way that they can adapt to changing user needs and image types. Besides automatic feature adaptation it also employs a weight based relevance feedback component.

The ImageRover system, developed at Boston University [236], combines colour and texture based descriptions obtained from fixed sub regions of images. Colour histograms are computed in the $L^*u^*v^*$ space whereas texture features are obtained with the help of steerable pyramids. Additionally feature vector dimensions are reduced with a PCA and stored via kD trees. The ImageRover system is designed for the WWW and equipped with a relevance feedback mechanism.

Blobworld, a system developed at the UC Berkeley [33], employs region based features computed on images segmented with the help of the expectation maximisation algorithm. Colour histograms are calculated on the $L^*a^*b^*$ space while texture is described by means of the mean contrast and anisotropy over a region.

The PicSOM image browsing system, developed at the Helsinki University of Technology [136], is based on a tree structured version of self organising maps (SOM). Consequently, a hierarchical image organisation is obtained, with "similar" images closer to each other. PicSOM employs average RGB colour vectors, MPEG-7 features as well as Fourier based shape descriptors. Moreover, relevance feedback is realised in an iterative manner according to user selections.

SurfImage is developed by Inria, France [188], and incorporates a variety of low and high level features. Basic descriptors such as RGB colour histogram, edge histogram and co-occurrence matrices are employed among higher level features such as eigenimages and deformable intensity surfaces. Multiple similarity metrics are implemented and weighted queries are also put to user disposal. A relevance feedback mechanism is additionally available, taking into account both positive and negative judgements.

The Informedia project [284] is one of the more recent video and film indexing systems, developed at Carnegie Mellon University. Along with the Marvel system, it constitutes one of the pioneering efforts concentrating on semantic descriptions.

Finally, the Viper system (Visual Information Processing for Enhanced Retrieval) is developed by the university of Geneva [254]. It employs more than 80.000 simple low level colour and spatial frequency features at several scales, chosen with the purpose of approximating the behaviour of the HVS. The system is additionally equipped with a relevance feedback support and can be tested online at [282].

Web based CBIR systems: WWW oriented image retrieval systems deserve special attention as they are faced with a greater challenge. The WWW represents one of the biggest and most heterogeneous visual repositories. Several application possibilities have emerged in this area (e. g. crime prevention, copyright protection, entertainment, etc), and subsequently various efforts were made in order to create effective retrieval systems. An survey on web based CBIR systems can be found in Ref. [133]. The availability of image captions and meta-data is invaluable in this case, as they aid considerably the indexing process. An example based on a web crawler is given in [230]. Systems using more visual content oriented techniques, in combination with image captions, include iFind [299] and Cortina [217].

5.4 Conclusion

In this chapter we have presented an overview of the rapidly maturing field of content-based image retrieval. We have elaborated on each component of the standard CBIR system architecture and reviewed the related past and current achievements. Moreover, a non exhaustive list of popular and classic CBIR systems has also been provided, highlighting their primary properties. Although it is widely recognised that the majority of the so far obtained results rely on low level visual features, the shift towards more sophisticated architectures is certainly under way. As the field has matured compared to its early years, the main issues have been identified and efforts have been allocated accordingly. Furthermore, given the huge demand for an effective retrieval system, the number of disciplines participating in the field of CBIR is increasing as is the diversity of the proposed methodologies. An insight into the current requirements from image retrieval systems and the rising trends have also been discussed.

In conclusion, although it is not yet known if it is possible to fully simulate the image understanding capabilities of the HVS, it is widely agreed upon that in order to obtain a successful CBIR system, the collaboration of multiple disciplines, with focus on each aspect of the standard CBIR system architecture, is imperative. Having presented the main points of the CBIR field, the next chapter will focus on the contribution of colour morphology to it.

Chapter 6

Morphological content descriptors

6.1 Introduction

The first foundations of mathematical morphology (MM) were laid down during the sixties, within the context of stereology related projects aiming to describe iron ore properties [239] and porous media [170]. Hence visual content description can be considered as one of the initial goals, that guided the research efforts leading eventually to the development of the entire morphological framework as it is known today.

Since that first period the potential of MM for feature extraction and content description has been demonstrated in various fields, ranging from biomedical applications [187] to remote sensing [145]. The topic of general purpose content based image retrieval (CBIR) however, with the exception of highly application specific cases (e. g. hematological cell classification [6]), has been left relatively unexplored. At the end of the inception years of CBIR, this has started to change. Rather direct applications of classical morphological operators, such as pattern spectra [70], combinations with already known descriptors (Gabor filters and watershed based segmentation [215]) as well as novel morphological approaches [4] are becoming more frequent. Nevertheless, to the best of the authors' knowledge no extensive study has yet been carried out.

Besides the relative popularity of CBIR in the image processing community, the motivation for considering morphological operators in order to tackle this problem, is probably due to the inherent capacity of this framework to exploit spatial pixel relationships. Based on this concept, the main goal of this chapter is to empirically explore the contribution potential of colour morphological approaches based on the results of Chapter 4 to two visual feature families, namely colour and texture, in the particular context of CBIR. Even though according to Chapter 5, shape is also an intrinsic visual property, we do not elaborate any further on it due to the following reason: as shape is a property describing the border configuration of objects, it requires an automatic, precise and reliable segmentation step, a tool which we lack. Effective morphology based colour segmentation approaches have been of course reported [12, 13, 244], however their reliability has not been deemed sufficient in the general case. Consequently, we propose here a number of different morphology based colour and texture descriptors, that are compared against some of the known alternatives presented in Sec. 5.2.3 in a series of comparative tests, using various image databases.

We proceed by studying each feature group independently. Specifically, we present three global colour descriptors (Sec. 6.2), making use of granulometries and multi-resolution histograms in combination with colour levelings and the watershed transform. We further introduce a texture descriptor (Sec. 6.3), combining the characteristics of morphological covariance and granulometries through the use of multiple structuring element variables. The tests that are carried out show that the proposed approaches provide competitive performances.

6.2 Colour

Colour is widely regarded as one of the most expressive visual features [165], and as such it has been extensively studied in the context of CBIR, thus leading to a rich variety of descriptors [256]. Similarly to texture and shape, it is relatively hard to establish the exact description of colour content. One has to ask, given a colour image, which notions describe the concept of colour within it. Considering the existing amount of work on this topic, and the multitude of colour descriptors employed in commercial and research systems, one can narrow down the type of information extracted from them to the following:

- The presence (or absence) of a particular colour (e. g. red and blue)
- The relative amount in which a particular colour is present (e. g. 80% red and 20% blue)
- The size distribution of each colour (e. g. one large patch of red and multiple small patches of blue)
- The relative spatial position of each colour with respect to others (e. g. red at left and blue at right)

The last of the aforementioned four notions is of use only in cases where the relative position of colour components is of semantic importance. For example in the case of beaches, the sand (brown and/or grey) is next to the sea (blue and green), above which is the sky (blue and/or grey). Of course there are also numerous other objects for which relative positions cannot be a priori determined, for instance accordions, books etc. Consequently, one can understand why the vast majority of existing work on this topic concentrates on the first three notions in order to describe colour content.

Any attempt on colour description however has to address certain issues first. These are the choice of colour space and quantisation levels, as well as the variations due to different illumination conditions. The choice of colour space is of vital importance, as it has a direct effect on channel-wise correlation, and consequently on the effectiveness of the resulting feature vectors. Although perceptual spaces (e. g. $L^*a^*b^*$, $L^*u^*v^*$) are considered to better model the human perception of colour, comparative studies have shown phenomenal colour spaces to outperform them in combination with several descriptors [161].

As the standard number of colours in everyday digital images is usually in the order of millions, colour specific approaches naturally lead to excessively large feature vectors, hence colour sub-quantisation (i. e. quantisation to fewer levels) is common practice. The exact number of levels along with their uniformity however, depend strongly on the chosen colour space. For instance while 4 colours per channel are preferred for RGB (i. e. 4-4-4, 64 colours), 16-4-4 is usually used with HSV [165, 191]. The question of choosing the optimal quantisation model and number of levels can be considered as a search for the optimal compromise between feature vector size and effectiveness.

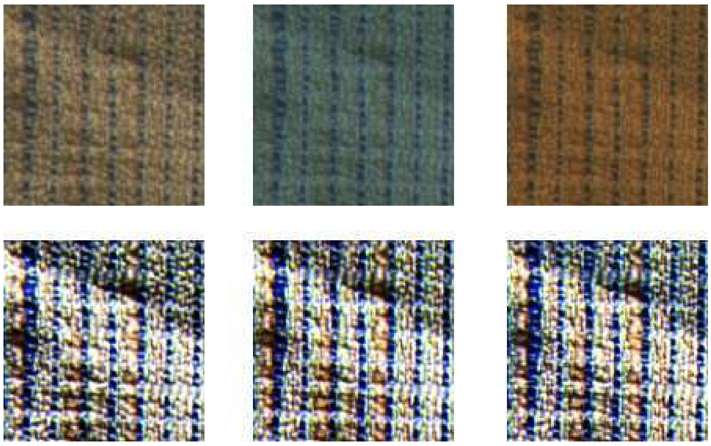


Figure 6.1: In the first row a texture example from the Outex14 texture database under three different illumination sources (Left to right, 2856K incandescent CIE A, 2300K horizon sunlight and 4000K fluorescent TL84) and in the second row, the same images after the application of a channelwise histogram equalisation in RGB.

Another significant obstacle in colour description is the one caused by colour variations due to different illumination conditions. Even with exactly the same object and viewing angle, changing only the illumination source may lead to totally different colour distributions. One can counter this effect, either by means of an appropriate transformation, like histogram equalisation (Fig. 6.1) [73] or by subquantising strongly the luminance component in order to reduce its influence [194].

Despite the absence of previous work on the morphological characterisation of colour distributions, we believe that morphological options possess the potential to contribute to all three stages of it, specifically preprocessing (e. g. image simplification), processing (e. g. feature extraction) and postprocessing (e. g. descriptor refinement). In the light of the aforementioned colour description principles, we present three global/holistic approaches for the description of colour based on morphological operators, that are subsequently compared against known alternatives.

6.2.1 Colour specific granulometry (CSG)

Since the first two requirements for describing colour content can be trivially obtained through an histogram, the main problem to address is finding a means to describe the spatial/size distribution of colours. To this end, an important part of the existing methods follows a colour specific approach, where each colour is described independently, either by means of an auto-correlogram or some other spatial histogram extension, each colour bin is associated additionally with a spatial description. That is why it has been decided here not to employ a vector processing strategy, and instead focus on each quantised colour separately using granulometry as our description tool. Besides, CBIR oriented colour granulometries have already been reported, such as the one by Angulo and Serra [10], where they are computed marginally in the RGB space and the one by Louverdis et al. [155], defined in HSV using a lexicographical ordering, which however ignores the hue component's periodicity.

Intuitively, we choose to associate each colour of the input with its corresponding granulometric curve (detailed in Sec. 6.3). Specifically, given a colour image $\mathbf{f}$ with possible colours in $\{c_i\}_{1\leq i\leq n}$, a grey-level image f_{c_i} for each c_i is computed, where every pixel $f_{c_i}(x,y)$ denotes the distance of $\mathbf{f}(x,y)$ to c_i:

$$\forall(x,y),\ f_{c_i}(x,y) = d(\mathbf{f}(x,y), c_i) \tag{6.2.1}$$

where d represents a colour metric suitable for the colour space under consideration, e. g. the Euclidean distance for perceptually uniform spaces. Next we compute the granulometric curve of f_{c_i} using closings by reconstruction. To explain, closing by reconstruction is a high level morphological operator based on geodesic erosions. Geodesic erosion $\varepsilon_g^{(1)}(f)$ and dually geodesic dilation $\delta_g^{(1)}(f)$ are operations involving two images, a marker image f and a mask g, where the first is processed conditionally to the second:

$$\varepsilon_g^{(1)}(f) = \varepsilon(f) \vee g \tag{6.2.2}$$

$$\delta_g^{(1)}(f) = \delta(f) \wedge g \tag{6.2.3}$$

where ε and δ are the erosion and dilation by the neighborhood of the origin. Both operators can be further applied successively as follows:

$$\psi_g^{(n+1)}(f) = \psi_g^{(1)}\left(\psi_g^{(n)}(f)\right) \tag{6.2.4}$$

hence by repeating them until stability (i times, $\psi_g^{(i)}(f) = \psi_g^{(i+1)}(f)$) one can realise respectively a reconstruction by erosion and a reconstruction by dilation:

$$R_g^{\varepsilon}(f) = \varepsilon_g^{(i)}(f) \tag{6.2.5}$$

$$R_g^{\delta}(f) = \delta_g^{(i)}(f) \tag{6.2.6}$$

Consequently one can define closing by reconstruction with a SE B:

$$\phi_{R,B}(f) = R_g^{\varepsilon}\left[\delta_B(f)\right] \tag{6.2.7}$$

which we use in combination with SE of various sizes, in order to compute the granulometric curve of f_{c_i} that describes colour c_i:

$$\mathrm{h}_f^k(c_i) = \mathrm{Vol}(\phi_{R,B_k}(f_{c_i}))/\,\mathrm{Vol}(f_{c_i}) \tag{6.2.8}$$

$$\forall c_i,\ \mathrm{h}_f(c_i) = \left\{\mathrm{h}_f^k(c_i) \mid k \in \{1,\ldots,m\}\right\} \tag{6.2.9}$$

where B_k is the SE of size k to be used in the closing by reconstruction, and Vol denotes the image volume (i. e. sum of grey-level pixel values of each f_{c_i}). Consequently, the final feature vector is formed by the concatenation of all $h_f(c_i)$, leading to a descriptor of size $n \times m$. The reason for using a closing instead of an opening is because the regions representing colours similar to the one under consideration will be at minimal distance, hence dark, whereas the choice of a reconstruction based operator is justified by the reports concerning the improvement of performance that it brings in the context of content description [267].

Moreover, granulometries, similarly to correlograms, are invariant with respect to spatial position [291], and provide a description of the size distribution of their content. Applying them directly on the input image results in a fairly effective shape descriptor [70].

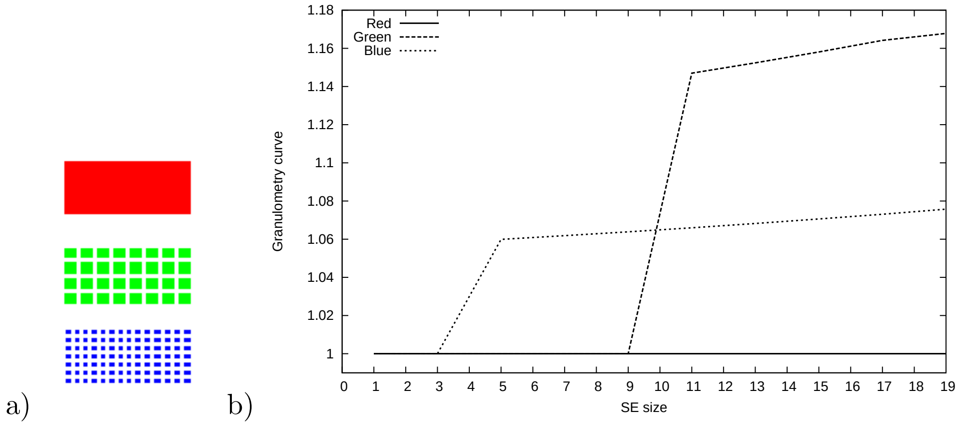


Figure 6.2: a) The input image b) the granulometric curves coresponding to red, green and blue, obtained by processing the distance image of (a) computed using an Euclidean distance in RGB.

By applying them "marginally" for each possible colour, one can distinguish for instance a connected "large" patch of colour c_i, from a number of smaller components of the same colour. As to its similarity comparison, we find it pertinent to use an approach similar to the relative L_1 distance based measure of correlograms, Eq. (5.2.8):

$$d_{CSG}(\mathbf{f}, \mathbf{f}') = \sum_{1 \leq i \leq n, 1 \leq k \leq m} \frac{|h_f^k(c_i) - h_{f'}^k(c_i)|}{1 + h_f^k(c_i) + h_{f'}^k(c_i))} \tag{6.2.10}$$

An illustrative example of CSG is given in Fig. 6.2, where the curves obtained for three colours of different size distribution are shown. Possible extensions to the definition of Eq. (6.2.9) could include using multivariate granulometries (e. g. size-shape), as well as higher order statistical measures instead of the volume.

6.2.2 Multi-resolution histograms in the leveling scale-space (MHL)

Colour histograms remain simple yet effective tools since the early days of CBIR [257], with positive properties such as invariance to geometric transformations. As they are capable of providing the first two elements required for the description of colour, their foremost drawback is the lack of any spatial information. To this end, several extensions have been proposed [219, 256]; among which one particular approach is using multiple resolutions, that has proven itself to be able to capture effectively colour content [92, 141]. Multi-resolution decomposition has been achieved by means of Gaussian filters [92], Gabor filters [120] as well as wavelets [212]. For an in-depth study on this topic the reader can consult Ref. [92]. Morphological scale spaces however have not yet been used to this end.

Scale-spaces based on morphological levelings have been studied in particular by Meyer and Maragos [175], where the useful properties of levelings, such as contour preservation and invariance to translation and rotation among others have been mentioned. This formulation later on has been used for optimal scale selection by exploiting the property of maxima propagation within levelings [269]. Here however, we propose to employ this scale space for the description of colour distribution by means of multi-resolution histograms. Before proceeding any further, let us recall the relative definitions. Levelings are powerful morphological operators representing a subclass of connected operators, that work on a reference image $\mathbf{f}$ and a marker image $\mathbf{g}$. The marker image is modified in such a way that it becomes the leveling $\mathbf{g}' = \lambda(\mathbf{f}, \mathbf{g})$ of $\mathbf{f}$. From an implementational point of view, the expression:

$$\mathbf{g}' = [\mathbf{f} \wedge \boldsymbol{\delta}(\mathbf{g})] \vee \boldsymbol{\varepsilon}(\mathbf{g}) \tag{6.2.11}$$

is iterated until idempotence, although more efficient algorithms also exist [174].

At this point two issues need to be addressed, first the extension of levelings to colour images, as well as the formulation of a scale-space. The first has been studied by Gomila and Meyer [83] and by Angulo and Serra in Ref. [11], where different orderings and colour spaces have been tested. It consists in replacing the scalar morphological as well as binary operators of Eq. (6.2.11) by their vector versions. As to the second, multiple methods are possible such as using a family of extensive and anti-extensive operators within Eq. (6.2.11) or a family of marker images $\{\mathbf{g}_i\}$ that lead to progressively simplified images:

$$\text{For resolution } i \in \{1, \ldots, n\}, \;\; \mathbf{g}'_0 = \mathbf{f} \text{ and } \mathbf{g}'_i = \lambda(\mathbf{f}, \mathbf{g}_i) \tag{6.2.12}$$

Of course, a variety of filters may be used in order to obtain the marker images, including *alternating sequantial filters* (ASF), openings by reconstruction (RO), Gaussian or even anisotropic diffusion filters, that according to the study in Ref. [128] provide the best quantitative performance in terms of noise reduction capacity. An example of successive colour levelings using multiple marker images obtained through opening by reconstruction is given in Fig. 6.3.

Figure 6.3: Three vectorly leveled versions of the Lenna image using Eq. 6.2.11, with marker images obtained using openings by reconstruction along with square SE of size 11, 23 and 35 respectively from left to right (This image is best viewed in colour).

Given n resolutions, one can then compute the colour histogram of size m of each g'_i and concatenate them in order to form the feature vector of size $n \times m$. Fig. 6.4 shows the histograms corresponding to the leveled images of Fig. 6.3, where the inter-resolution differences can be observed. Moreover, the histogram differences may also be used, hence leading to a $(n-1) \times m$ sized vector. As to the similarity measure, while the histogram intersection is of course the standard choice, a weighting factor may also be added, following the principle of a *pyramid match* [87]. In detail, since highly simplified images are bound to be more alike than the lower resolutions where more image detail is preserved, it makes sense to weight the corresponding histogram intersections. For instance, given $H(\mathbf{f}) = \{H_i(\mathbf{f})\}_{0 \leq i \leq n}$

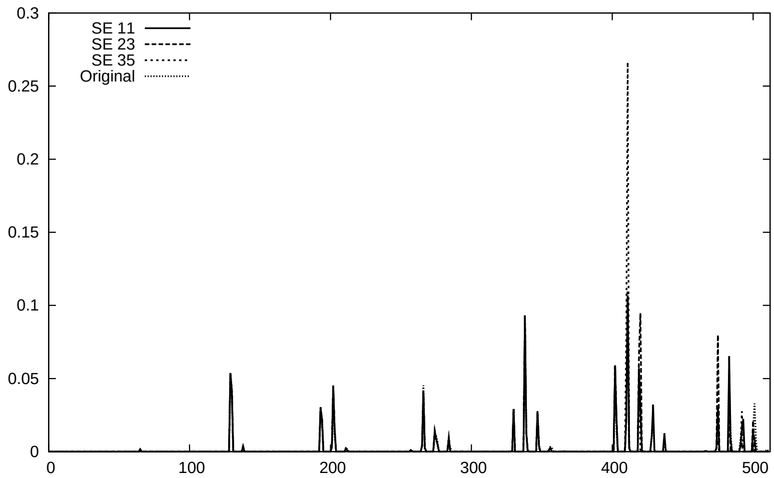


Figure 6.4: The 512-bin RGB colour histograms of the leveling example shown in Fig. 6.3.

and $H(\mathbf{h}) = \{H_i(\mathbf{h})\}_{0 \leq i \leq n}$, the multi-resolution histogram families for two images $\mathbf{f}$ and $\mathbf{h}$ respectively, their weighted distance d_w would become:

$$d_w(H(\mathbf{f}), H(\mathbf{h})) \; = \; \sum_{i=0}^{n} \frac{1}{2^i} \sum_{j=1}^{m} \min \{H_{ij}(\mathbf{f}), H_{ij}(\mathbf{h})\} \qquad (6.2.13)$$

where H_{ij} denotes the j^{th} histogram bin of resolution i. Further refinement may be achieved by modifying the bin width at each resolution.

6.2.3 Multi-resolution histograms based on watersheds (MHW)

Another morphological tool that may be used in order to boost the spatial sensitivity of the standard colour histogram is the watershed transform. This powerful operator represents the foremost morphological approach to the problem of image segmentation. Although fully automatic, it is relatively sensitive to image noise, hence leading often to oversegmented results unless some form of smoothing is applied first. Nevertheless, the regions that are formed represent uniform areas in terms of spectral content or "flat" in topological terms. Here we propose to use this operator in order to produce the different image resolutions that are to be used in order to construct a multi-resolution histogram.

In particular, given a colour image $\mathbf{f}$, it is first smoothed relatively to resolution i and then segmented by means of the watershed transform, which leads usually to several regions. Then, the pixels of each region are associated with the average colour of that region. Consequently one can at this point compute the standard histogram taking into account the newly formed regions. Besides, using multiple resolutions means that the initial image will be further smoothed hence leading to the formation of even larger flat regions, and thus we obtain a multi-resolution representation of the input the histograms of which capture indirectly the spatial organisation of its content. This procedure is further illustrated in Fig. 6.5. Consequently the end feature vector is formed by the concatenation of the histograms of each resolution, thus leading to a size of $n \times m$, where n is the number of resolutions and m the number of bins of each histogram. As a similarity measure, we consider the pyramid match of Eq. (6.2.13) as adequate.

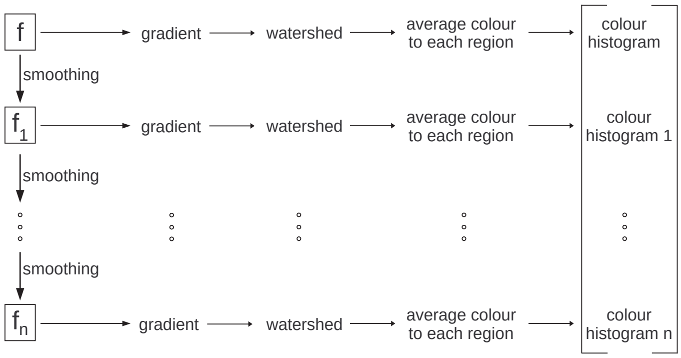


Figure 6.5: An illustration of the steps required for forming the watershed transform based multi-resolution histograms.

Figure 6.6: The results obtained after the watershed computation on the levelings of Fig. 6.3, where pixel values have been replaced by the average colour of each basin.

Although it appears as a straightforward approach, its implementation is hindered however by certain elements. First, there is the issue of a morphological colour image segmentation [12]. The watershed transform is most often applied on a grey-level input representing the topographic relief of the image under consideration, usually its gradient, that is why an effective means of computing the colour variations is necessary. To this end, we choose to combine the colour channels by means of a channelwise maximum of marginal gradients:

$$\forall\, l, s, h \in [0,1]^3 \;\; \rho_{LSH}(l,s,h) = \max\{\rho(l), \rho(s), \rho_H(h,s)\} \tag{6.2.14}$$

where $\rho = f - \varepsilon(f)$ is the standard internal morphological gradient, chosen instead of its alternatives based on experimental observations. Although the components of the polar colour spaces are highly intuitive, their combination is relatively problematic. In particular, hue is of no importance if saturation is "low", while the bi-conic shape of the colour space assures that no high saturation levels exist, if luminance is not "high enough". Hence the hue gradient needs to be weighted with a coefficient that has a strong output only when both compared saturation values are "sufficiently high". Besides, on the contrary of the other two components, it has been observed that the max – min form provides visually superior results

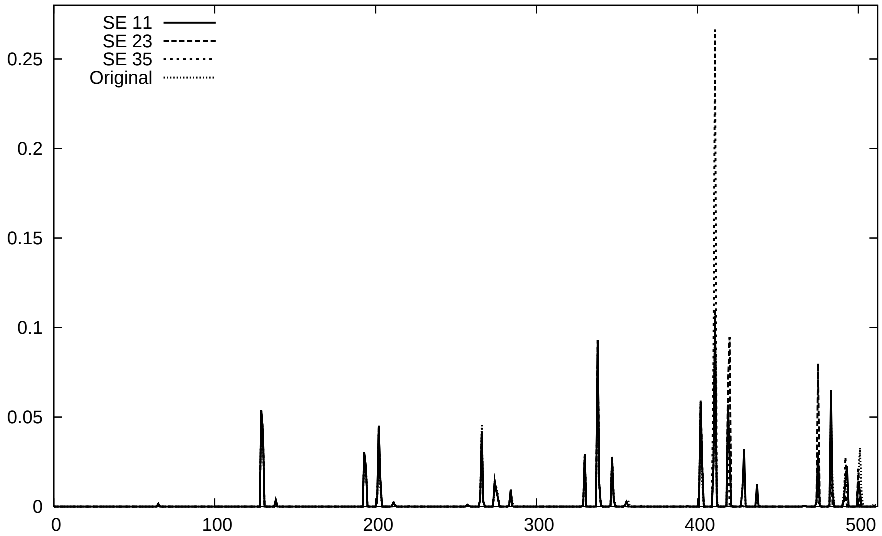


Figure 6.7: The histograms of the images of Fig. 6.6.

with the hue, hence leading us to use the following expression:

$$\rho_H(h, s) = \max_{i \in B} \{j(s, s_i) \times h \div h_i\} - \min_{i \in B} \{j(s, s_i) \times h \div h_i\} \tag{6.2.15}$$

where B is the local 8-neighborhood, $\div$ the hue distance defined in Eq. (4.2.4) and $j(\cdot, \cdot)$ a double sigmoid controlling the transition from "low" to "high" saturation levels:

$$j(s_1, s_2) = \frac{1}{(1 + \exp(\alpha \times (s_1 - \beta))) \times (1 + \exp(\alpha \times (s_2 - \beta)))} \tag{6.2.16}$$

where the slope $\alpha = -10$ and the offset $\beta = 0.5$.

Moreover, one also has to choose a means to compute the progressively smoothed versions of the input image. Here, we consider colour levelings with this purpose, the marker images of which may be obtained with any of the options mentioned in Sec. 6.2.2. Following the flattening realised by levelings, the flat regions will be easily detected by the watershed transform, that will also detect as regions all basin shaped areas that have not been flattened by levelings. Whether this nuance has a positive or negative effect on the operator's colour content description capacity is a question to be answered experimentally in Sec. 6.2.4. As an example, the application of the watershed transform on the computed gradient leads to the result shown in Fig. 6.6, whereas Fig. 6.7 shows their histograms, where inter-resolution differences are present though hard to observe.

6.2.4 Experimental evaluation

Setup

The retrieval tests that have been carried out have served multiple purposes. First they have been used in order to verify the practical interest of the presented morphological colour descriptors, as well as to determine the optimal colour space and quantisation couple for this task. To this end, a total of 3000 images from three different sources have been used

[52, 88, 84]. This set contained 16 semantic categories of 100 images each, along with 1400 bulk images, in order to simulate real-life retrieval from an unorganised database. Examples from each category are given in Fig. 6.8. Furthermore, in order to obtain more reliable estimates, each image of the 16 categories served as the query subject in turn, making a total of 1600 queries. From the results of which the precision vs recall charts have been computed, as well as ANMRR values (the MPEG-7 retrieval effectiveness measure), that provides an overall rating of performance. It should be additionally noted that for the charts the first retrieved image was always ignored as it represents the query image itself. In our tests, the proposed morphological colour descriptors have been compared against the standard *histogram*, the *auto-correlogram* (AC) [115], the superiority of which with respect to the histogram has been shown by multiple comparative studies [161, 194], the *colour structure descriptor* (CSD) [165] of the MPEG-7 standard and colour distribution entropy (CDE) [256].

Moreover, a question that arises at this point is "why use a non standard image database?". Indeed, our initial intent was in the opposite direction, however it became soon clear that contemporary image collections, such as Caltech-256 [88] are rather unsuitable for this particular testing suite, in the sense that they require descriptors capable of taking into account multiple visual dimensions (e. g. colour, shape, etc) simultaneously, thus rendering them inappropriate for the present case where we wish to focus on colour, and colour only. Of course this also represents a limit for this series of tests, as the results rest valid for collections with a certain level of colour separability among categories.

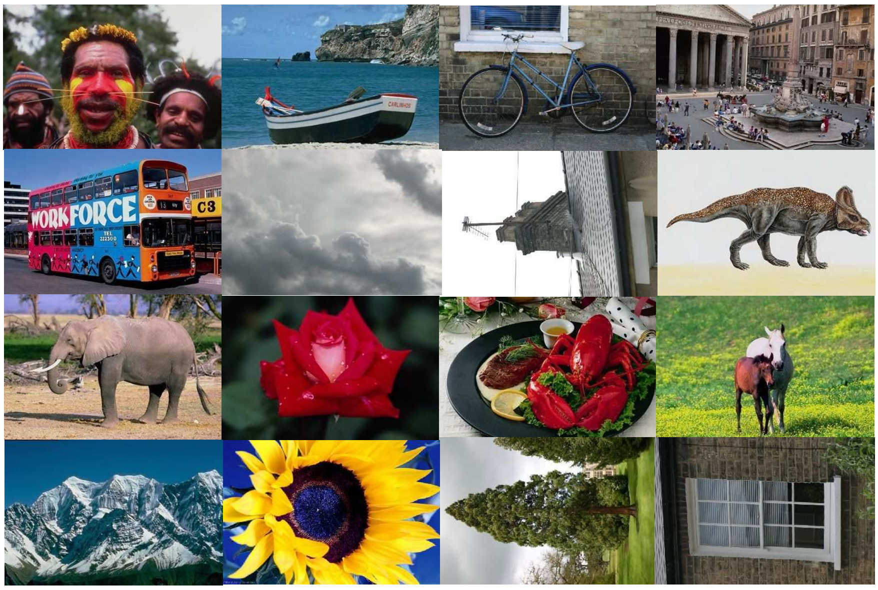


Figure 6.8: Representative category examples of the image database used for colour descriptor comparisons. From top-left to bottom-right: Africa, beach, bicycles, building, bus, clouds, chimneys, dinosaur, elephant, flower, food, horse, mountain, sunflower, tree and window.

Test results

Colour space and quantisation: According to previous comparative studies [161] and the MPEG-7 standard [165], polar colour spaces in general outperform most of their alternatives in the context of content description, in combination with different quantisation models. Considering however the design problems of these spaces (Sec. 2.7.1), and the alternative formulations that have been proposed, it has been considered appropriate to test their effectiveness in order to determine the most suitable colour space along with its quantisation model for the image database under consideration.

To this end, we have employed colour histograms along with RGB, $L^*a^*b^*$, HSV, and LSH with 3 different uniform quantisation models, namely 8 (8-8-8), 6 (6-6-6) and 4 (4-4-4) colours per channel. The resulting ANMRR values are given in Table 6.1. Judging from the obtained results, there are primarily two remarks to be made: first, the general superiority of LSH with respect to its counterparts, a result which confirms previous studies [194]. And additionally, one can observe the relatively light influence of histogram bin numbers. As a matter of fact, even with eight times less colours, the performance of the histograms decreases only by 8 to 10%.

Colour spaces	8-8-8 (512)	6-6-6 (216)	4-4-4 (64)
RGB	0.4993	0.5042	0.5478
$L^*a^*b^*$	0.5544	0.5718	0.6127
HSV	0.4912	0.4998	0.5518
LSH	0.4759	0.4824	0.5158

Table 6.1: ANMRR values for different quantisation models (histogram bins) and colour spaces.

Based on the performance of the two polar colour spaces, from this point on we focus on the optimisation of their arguments. For instance, which brightness expression is most adequate for this task? Value (maximum of R, G, B), intensity (average of R, G, B), or luminance (perceptually weighted combination of R, G and B)? Table 6.2 presents the results obtained using a 256 bin brightness histogram with the aforementioned expressions. According to the obtained results, the relatively simple expression of red green blue average, provides an overall better performance with this set.

Value	Intensity	Luminance
0.707	0.6843	0.7086

Table 6.2: ANMRR values for different expressions of brightness using 256 bin brightness histograms.

Proceeding to the other two channels, as far as saturation is concerned, we test the brightness dependent version of HSV against the brightness independent definition of LSH, by means of a 256 bin saturation histogram. Table 6.3 shows the obtained ANMRR values. According to the results, brightness independence aids saturation's retrieval performance.

As to the hue component, we choose the implementation of LSH, and test the effect of using a saturation weighted hue histogram against a "raw" hue histogram, where the hue values of achromatic pixels take place as well. Table 6.4 shows the obtained results. The weighting factor has indeed a positive effect.

In the light of these results, we opt for the LSH colour space, equipped with the R, G, B average as brightness expression, and its hue weighted by its saturation. Furthermore,

Luminance dependent	Independent
0.7053	0.6655

Table 6.3: ANMRR values for different expressions of saturation using 256 bin saturation histograms.

Saturation weighted	unweighted
0.5892	0.6275

Table 6.4: ANMRR values for saturation weighted and unweighted 256 bin hue histograms.

since the hue histogram is evidently more pertinent with respect to brightness and saturation, similarly to the general tendency [165, 191], non-uniform quantisation of LSH is tested next. Besides, allowing a lower number of levels for brightness, has the additional advantage of limiting the effect of illumination variations. But the question is, how exactly should one quantise this space? Common models include 4-4-16 [165], 3-3-12 [194] as well as 3-3-9 [256]. Moreover one has also the option of quantising the hue circle non uniformly, according to [143], which is based on the fact that human colour perception has a varying capacity of hue distinction. For instance, the average person can distinguish more hues of red than blue. This principle is illustrated in Fig. 6.9, and the quantisation option is tested as 3-3-7 (NU). Table 6.5 summarises the results obtained from various quantisation models. Although in general the exact configuration has relatively small effect on the retrieval performance, as a practical compromise between colour number (i. e. feature size) and effectiveness, the non uniform 3-3-7 stands out. Hence it will be subsequently used as the default quantisation scheme, unless stated otherwise.

Quantisation model	Colour number	ANMRR
3-3-7 (NU)	63	0.4908
3-3-8	72	0.5045
4-4-8	128	0.4912
3-3-12	108	0.4813
3-3-16	144	0.4800
4-4-12	192	0.4729
4-4-16	256	0.4717

Table 6.5: ANMRR values for various quantisation models. NU stands for non-uniform.

Descriptors: Having chosen the LSH colour space with a non uniform sub-quantisation of type 3-3-7, we now continue to compare the colour descriptors presented in Sec. 6.2 against some of the classical methods of colour content description. As far as CSG is concerned, the projection of each quantised colour is computed using the distance between the colour under consideration and the original unquantised image. The distance expression is set to a weighted combination of hue and luminance difference:

$$\forall c_1 = \{h_1, s_1, l_1\}, c_2 = \{h_2, s_2, l_2\}, d(c_1, c_2) = (h_1 \div h_2) \times j(s_1, s_2) + |l_1 - l_2| \times (1 - j(s_1, s_2)) \tag{6.2.17}$$

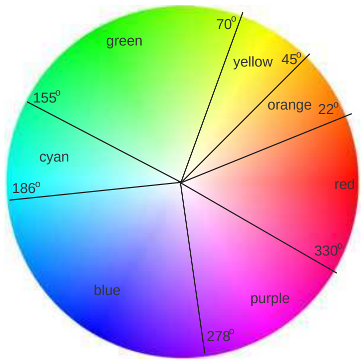


Figure 6.9: Non uniform hue circle quantisation.

where all colour components have been normalised to [0,1], the weight $j(\cdot,\cdot)$ is given in Eq. (6.2.16) and $\div$ represents the standard hue distance defined in Eq. (4.2.4). Moreover, the granulometries have been computed using square shaped SE with 4 sizes, 1, 5, 9 and 13 pixels wide using the efficient algorithm of Ref. [281]. Thus, with $3 \times 3 \times 7 = 63$ colours, it leads to a feature vector of length $4 \times 63 = 252$. As to MHL, based on empirical observations it has been decided to use openings by reconstruction in order to realise the levelings, in 4 different resolutions, with the same SE options as CSG. Levelings of the same setup have been also used along with MHW, prior to applying the watersheds. Consequently, all three operators lead to identical feature vector sizes.

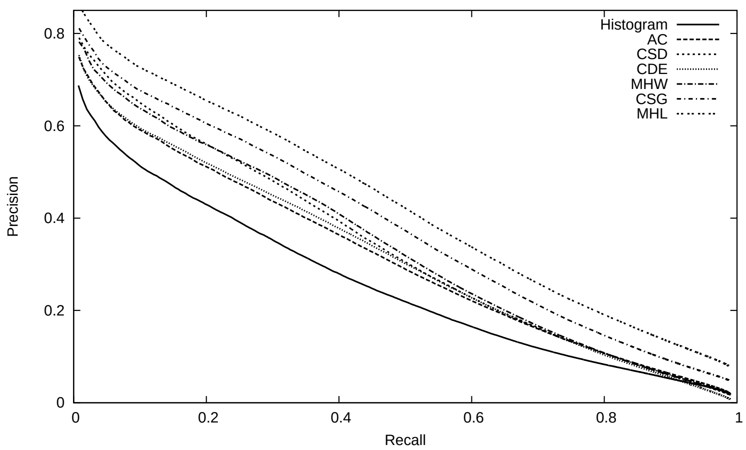


Figure 6.10: Precision vs recall chart of the tested colour descriptors.

On the other hand, as to their rivals, the standard colour histogram is computed in LSH using the sub-quantisation model 4-4-16 with histogram size 256. The AC is calculated with the same setup as in Ref. [194], where it is defined for the HSV colour space in combination with a sub-quantisation model of 12-3-3. Namely, with four distances (1,3,5,7), hence leading to a feature vector of length $108 \times 4 = 432$. For the CSD, we use the configuration presented in Ref. [165] but with the aforementioned colour space related choices instead of the HMMD space. Consequently it provides a feature vector of the same length as the standard histogram. In order to calculate the CDE we follow the definition given in Ref. [256], with the number of radius quantisation levels set to $N = 4$ which results in a description of length 252.

Histogram	AC	CSD	CDE	CSG	MHW	MHL
0.4717	0.4341	0.4132	0.4289	0.3935	0.4156	**0.3767**

Table 6.6: ANMRR values for the tested colour descriptors.

The precision vs recall charts resulting from the 1600 queries are shown in Fig. 6.10, whereas the ANMRR values are given in Table 6.6. As expected, the total lack of spatial information places the standard colour histogram last, while considerable improvements are achieved by both AC and CDE that provide likewise performances. Among the legacy descriptors, similarly to the study of Ojala et al. [191], the CSD gives the best retrieval results. As far as the proposed colour description operators are concerned, with the exception of MHW, they all lead to superior performances, especially in terms of precision. In particular, we can observe that the MHW has a similar output with CSD, while both CSG and MHL outperform their counterparts, with the latter providing the best results. We now proceed to test their robustness against noise.

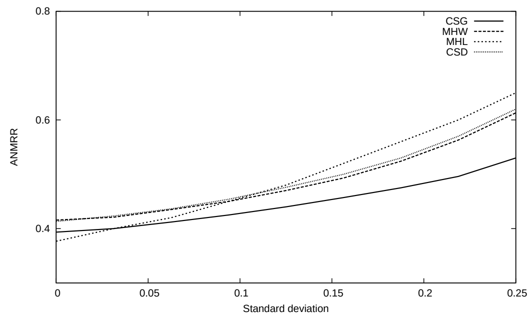


Figure 6.11: Chart of ANMRR values with respect to various noise deviation levels of the tested colour descriptors.

Robustness against noise and efficiency: To test the noise robustness of the proposed descriptors, we have corrupted the images of our database with various levels of zero-mean additive Gaussian noise, and measured their retrieval performances. In particular, the images have been contaminated with noise in the RGB colour space. The ANMRR values that have been obtained are shown in Fig. 6.11. This time they are compared only against CSD, as its noise robustness is known to be superior to its alternatives [165].

Histogram	AC	CSD	CDE	CSG	MHW	MHL
1	127	89	9	548	470	443

Table 6.7: Relative computation times of various descriptors to the standard colour histogram.

Judging from the obtained values, we can immediately remark the sensitivity of the previously outperforming MHL to noise levels, while MHW and CSD once more provide very similar outputs. CSG however exhibits a somewhat higher robustness. Further steps to improve their performance with corrupted image data could include the use of rank based morphological operators. As to their computation times, they are given in Table 6.7 with respect to that of a standard colour histogram. Generally speaking all three morphology based approaches are significantly slower than their counterparts, nevertheless they still possess a certain margin for further efficiency optimisation. Hence, one tends to conclude that they are unsuitable, at least in this form, for online retrieval systems.

In conclusion, the proposed morphological descriptors possess mixed properties, and while providing a retrieval performance as well as noise robustness comparable to contemporary approaches, they do have efficiency issues that need to be resolved if they are to be used along with online CBIR systems.

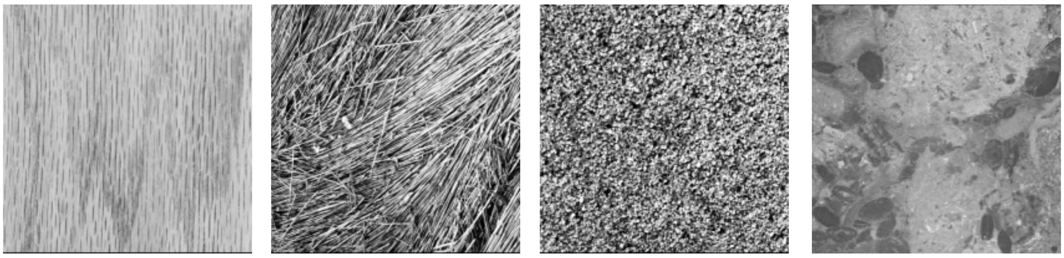


Figure 6.12: Texture examples from the Brodatz album [29], from left to right, strongly ordered, weakly ordered, disordered and compositional.

6.3 Texture

In this section, we start by recalling the basic texture categories along with their perceptual characteristics, and then the covariance and granulometry operators are elaborated, prior to presenting the proposed texture descriptor. According to the pioneering taxonomy work of Rao [220], textures can be classified with respect to their spatial distribution of details into four categories (Fig. 6.12).

- *Strongly ordered:* textures consisting of the repetitive placement of their primitive elements according to a particular set of rules.
- *Weakly ordered:* textures possessing a dominant local orientation, which can however vary at a global level.
- *Disordered:* textures lacking any repetitiveness and orientation, and usually described on the basis of their roughness.
- *Compositional:* textures that do not belong to any of the previous categories, and exhibit a combination of their characteristics.

In an effort to determine efficient features, capable of discriminating among the members of these categories, Rao and Lohse [221] have conducted psycho-physical experiments, and identified *regularity* (or *periodicity*), *directionality* and *complexity* as the most important perceptual texture characteristics, as far as human observers are concerned. With the subsequent work of Chetverikov [45] and Mojsilovic et al. [178], *overall colour* and *colour purity* have been added to this list. Given these texture properties, various operators have been defined in an effort to describe them (Sec. 5.2.3). As far as the morphological framework is concerned, it offers a variety of tools for texture characterisation, such as *granulometry*, *morphological covariance*, *orientation maps*, etc. The first two in particular have been employed successfully in a number of texture analysis applications [18, 78, 240, 252].

In particular, morphological covariance was initially proposed [170, 171, 240] as the equivalent in MM of the autocorrelation operator. The morphological covariance K of an image f, is defined as the volume Vol (i. e. sum of pixel values for grey-level images, and sum of the Euclidean norm of pixels in RGB for colour images), of the image, eroded by a pair of points $P_{2,v}$ separated by a vector v:

$$K(f; P_{2,v}) = \mathrm{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) \tag{6.3.1}$$

In practice, K is computed for varying lengths of $\mathbf{v}$, and most often the normalised version is used for measurements:

$$K^n(f) = \mathrm{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) / \mathrm{Vol}\left(f\right) \tag{6.3.2}$$

In the light of the aforementioned perceptual properties of textures, given the resulting uni-dimensional covariance series, one can gain insight into the structure of a given image [252]. In particular, the periodic nature of covariance is strongly related to that of its input. Furthermore, the period of periodic textures can easily be determined by the distance between the repeated peaks, that appear at multiples of the sought period; whereas the size of the periodic pattern can be quantified by means of the width of the peaks. In other words, their sharpness is directly proportional to the thinness of the texture patterns appearing in the input. Likewise, the initial slope at the origin provides an indication of the coarseness, with quick drop-off corresponding to coarse textures. In order to obtain additional information on the directionality of f, one can plot against not only different lengths of $\mathbf{v}$, but orientations as well [173].

Granulometry on the other hand [170, 171] as a term belongs to the field of materials science, where the granularity of materials is determined by passing them through sieves. Using the same principle, this operator consists in studying the amount of image detail removed by applying morphological openings γ_λ and/or closings ϕ_λ of increasing size λ. The volumes of the opened (or closed) images are then plotted against λ, or more usually their discrete derivative $\mathrm{Vol}(\gamma_\lambda - \gamma_{\lambda+1})$, i. e. *pattern spectrum.* The normalised version of the operator can be written as:

$$G^n(f) = \mathrm{Vol}\left(\gamma_\lambda(f)\right) / \mathrm{Vol}\left(f\right) \tag{6.3.3}$$

For unbiased measurements, the volume computation may be reduced only to the area affected by the operator. As a featuring tool, granulometry provides information on the shape and size of ordered textures, and regularity of disordered textures [23, 252].

6.3.1 Combining granulometry and covariance

Indeed, considering the aforementioned fundamental perceptual texture properties, morphological covariance and granulometry provide invaluable, yet complementary information on their input. More precisely, covariance extracts a feature vector containing information on periodicity and directionality, whereas granulometry concentrates rather on the granularity of its input. Consequently both are necessary in the general case for an efficient texture description. However, their combination is rather ambiguous, as it can be realised in a variety

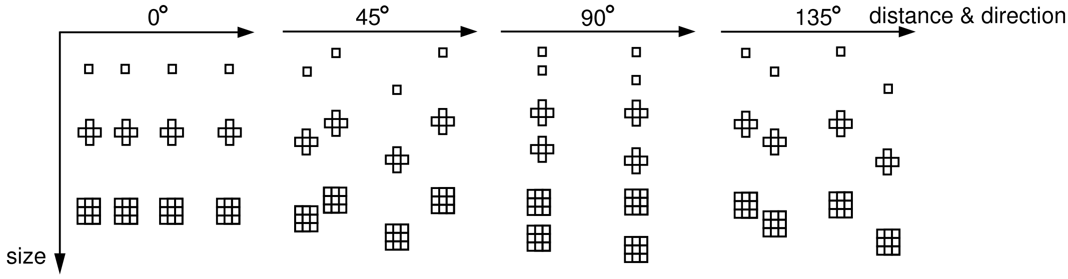


Figure 6.13: Illustration of structuring element pair variations, with respect to size, direction and distance.

of ways. The obvious method, is to calculate independently each feature vector and then employ their concatenation. We propose here an alternative way, which consists in unifying the two operators' functionalities by varying in parallel multiple SE properties. Moreover the use of multivariate granulometry and covariance has already been reported, specifically, in the form of combined SE direction and distance [101], as far as covariance is concerned, and shape and size combination with granulometry [267].

We choose to implement with this purpose a combination of SE size, direction and distance (Fig. 6.13). We further replace the erosion (ε) operator of covariance, Eq. (6.3.2) with an opening (γ). A connected alternative operator to γ (e. g. opening by reconstruction) could also be used for efficiency reasons, but at this early stage of operator design we have rather focused on its feasibility and left optimisation as a future priority task. Of course, on the contrary of granulometry it is also necessary to employ SE pairs, so that periodicity information may be extracted. Hence the following hybrid expression is obtained:

$$GK^n(f) = \mathrm{Vol}\left(\gamma_{P_{\lambda,v}}(f)\right) / \mathrm{Vol}\,(f) \tag{6.3.4}$$

where $P_{\lambda,v}$ denotes a pair of SE of size λ separated by a vector $\mathbf{v}$. Since $\mathbf{v}$ is the vector among the centers of the SE of the pair, naturally its length must be superior to λ to prevent overlapping ($\|\mathbf{v}\| > \lambda$). However, it should be noted that as the sieving principle of multiple morphological openings is satisfied if, and only if the SE is a compact convex set [171], this combination no longer qualifies as a granulometry. In practice, only the four basic directions $(0°, 45°, 90°, 135°)$ are of importance, thus it was chosen to integrate directional variation with distance as shown in Fig. 6.13. Of course, in case directionality becomes particularly significant one can always separate it as an additional dimension representing a finer distinction of directions, or even add one more dimension for shape distributions, where different SE shapes (e. g. disc approximation, square, lines, etc) are also employed along with direction, size and distance.

Fig. 6.14 presents the plots of the resulting features matrices, as applied to the strongly ordered and disordered textures of Fig. 6.12. Although their size distributions are rather similar, their directionality and periodicity are clearly distinct. Furthermore, as far as classification is concerned, feature vector or matrix size is of primary importance, since redundant information may eventually be present and disrupt the overall process. Even with the moderate sizes used in practice, the resulting feature set can easily become excessively large. That is why dimension reduction techniques, such as principal component analysis (PCA), could become necessary.

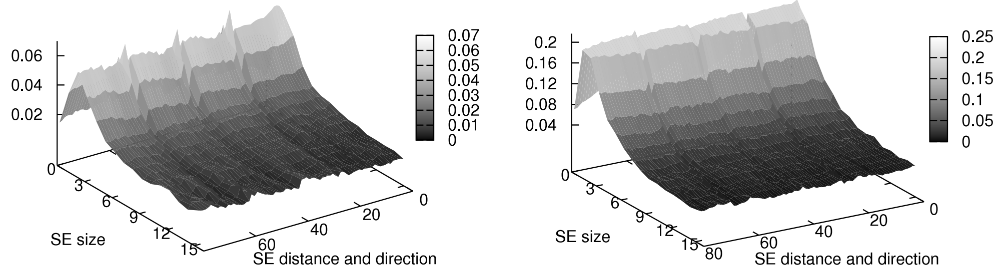


Figure 6.14: Plot of the feature matrices resulting from the application of Eq. (6.3.4) on the strongly ordered (left) and disordered (right) texture of Fig. 6.12.

6.3.2 Experimental evaluation

The proposed combination of granulometry and covariance has been tested using the colour textures of Outex13 (Sec. 3.4.2) [191]. We have compared four different feature extraction schemes with both grey-level and colour images, against the variogram operator, defined in Eq. (5.2.27). Specifically, we test features computed using a granulometry (Granulometry), Eq. (6.3.3), morphological covariance (Covariance), Eq. (6.3.2), their concatenation (Concatenated) and finally their proposed combined form (Combined), Eq. (6.3.4). More precisely, for granulometry square shaped SE have been employed, where a SE of size k has a side of $2k+1$ pixels, and k varied from 1 to 30 in steps of size 2. As to covariance, the four basic directions have been used $(0°, 45°, 90°, 135°)$ in combination with distances varying from 1 to 20. For their concatenated as well as proposed combined form, the same arguments were in place (with the distance values denoting the distance between the edges of the SE, not from their centers). The 80×15 feature matrix that has resulted from the combination option, has been reduced into a matrix of size 80×2 by means of a PCA transform and preserving only the first two dimensions. For grey-level computations the brightness component of LSH has been used. As far as the variogram is concerned, we have followed the setup of Ref. [101], so they have been computed in four directions $\alpha = 0°, 45°, 90°, 135°$, and for distances $q = 1, 2, \ldots, 50$. Thus leading by concatenation to a feature vector of length 200. The image set has been classified using a kNN classifier with $k = 1$ and the Euclidean distance as a similarity metric.

Features	Grey-level	Colour
Granulometry (Gra)	67.53	68.78
Variogram (Var)	75.56	76.53
Covariance (Cov)	73.82	76.92
Concatenated (Conca)	77.75	79.93
Combined (Combo)	**83.53**	**85.53**

Table 6.8: Classification accuracies (%) for the Outex 13 textures.

The classification accuracies, computed as the fraction formed by the number of successful classifications divided by the total number of subjects, are given in Table 6.8. Globally, one can immediately remark the positive, though comparably to intensity, small effect of using colour information. A result which asserts the auxiliary role of colour in texture recognition. Moreover, covariance systematically outperforms granulometry, using larger feature vectors, hence indicating the higher pertinence of periodicity and directionality with this database, compared to granularity. Additionally, the variogram, as a generalisation of covariance leads to similar results. The combination of the two operators (i. e. granulometry and covariance) on the other hand, by means of a concatenation improves the accuracy levels, while the proposed hybrid operator provides the overall best results, both with colour as well as grey-level images.

6.4 Conclusion

Following the presentation of various morphological colour image analysis approaches in the first part of this thesis, this chapter has focused on their use in the context of image content description. In particular we have concentrated on the notions of colour and texture at a global level, since no fully automatic as well as reliable segmentation approach was available at the time of study, that would have otherwise made it possible to use local descriptors. Furthermore, the task of shape description has also been omitted due to the same reason, as well as not being a colour related concept. As to interest points, they are with no doubt one of the best performing content description approaches of the last years [176], and their combination or at least implementation with morphological means appears to be a promising direction [222]. Nevertheless, it has been preferred not to follow it at this early stage of our work and rather concentrate first on the contribution potential of MM to classic description methods.

As far as colour is concerned, its principle properties from a descriptive point of view have been determined, and three global descriptors have been proposed, focusing mainly on the integration of spatial information with spectral colour distribution through morphological means. Specifically, the proposed approaches include the colour specific granulometry, which provides a size distribution independently computed for each sub-quantised colour, and two multi-resolution histograms based on morphological levelings and watersheds. The tests that have followed, have provided results asserting their competitive performance with respect to known alternatives. Nevertheless, the relatively high computation time that is required prohibits their use with online retrieval systems, while also leaving a considerable margin for speed optimisation.

As to texture, similarly to colour, first the fundamental properties of textures have been summarised, prior to concentrating on the two main morphological texture characterisation tools: granulometry and covariance. Since both provide complementary information on their input, we have opted for a rather unorthodox approach of their combination which leads to a hybrid operator, using multiple SE variables. Nevertheless it does not qualify theoretically as a granulometry, due to the shape of its SE. The tests that have been realised in combination with the Outex13 texture database, have confirmed the practical interest of this combination, as well as its superior performance with respect to the colour variogram. However, the multitude of SE variables, while capturing multiple textural properties also leads to an oversized feature matrix, the reduction of which becomes mandatory, by means of approaches such as PCA.

Equipped with content descriptors for colour images, the next chapter will elaborate on a content-based image retrieval architecture.

Chapter 7

MIMAR: yet another CBIR system

7.1 Introduction

Following our efforts focused on developing morphological colour analysis approaches (Chapter 4), that have been then somewhat used in order to define morphological visual content descriptors (Chapter 6), we now combine those results with the purpose of obtaining a content-based image retrieval system that can satisfy practical needs. In other words, semantic sensibility from the retrieval engine's part is necessary, along with other requirements. It is undoubtedly a great challenge, with multiple design related limitations.

We start with an overview of these restrictions and define our specific architecture goals accordingly (Sec. 7.2). In particular, solutions are needed for providing a means of retrieval by semantic content, while employing an appropriate set of descriptors, since none of them can be considered as suitable for *all possible queries*. Furthermore, restrictions specific to this thesis are also present, such as a certain level of adaptability to particular image collection domains. Of course, one also has to consider efficiency related issues besides effectiveness, since long retrieval times would render practically useless even the most effective of systems.

Consequently, to address these issues we concentrate on a keyword based retrieval engine, which exploits the semantics brought into the system by the expert's choice of example and counter example images (Sec. 7.3). Moreover, additional flexibility is achieved by means of dynamic keyword-descriptor associations. The presence of feedback support also aids in providing interesting results, while through the use of keywords, besides offering an easy to use and familiar retrieval means to users, it becomes possible to achieve constant time retrieval thanks to offline annotation. The exact implementational choices are then detailed in Sec. 7.4, whereas Sec. 7.5 provides the results of a series of tests with the system illustrating its functions. Finally, Sec. 7.6 contains concluding remarks and perspectives.

7.2 Design goals and restrictions

What is it that one needs from a CBIR system? Ideally, it would consist of a retrieval engine along with an user interface, making it possible to realise effectively and efficiently queries of all three levels, independently from the image database under consideration (Sec. 5.2.4). The various and considerable difficulties, hindering the realisation of this ideal system have been elaborated in Chapter 5. In the present case, given the considerable, yet not outstanding capacity of the descriptors presented in the previous chapter, we take a rather pragmatic direction with the purpose of exploiting them in the best way possible that will allow us to at least approach the qualities of the aforementioned ideal system, while also profiting from various design elements in CBIR that have been developed so far (e. g. relevance feedback, dynamic descriptor allocation). In particular, the design goals and restrictions that will determine the final architecture are the following.

1. First, since user needs are in their majority unrelated with level one queries [66], e. g. "retrieve 70% red images", the practical interest of the end system is highly dependent on its capacity to realise queries of semantic nature. Consequently, providing at least a certain degree of this function is our main goal, as well as of the entire CBIR field.

2. Moreover, as each retrievable image family e. g. sunsets, cars, etc, has a certain amount of a priori information available, e. g. a sunset is in the sky, cars have wheels, it is desirable to be able to accommodate these additional pieces of information with the end of improving the end results. Specifically, according to the financial sponsors of this project (Oséo ANVAR), it is required to provide a system capable of not only functioning with general image databases, but of adapting itself to case specific situations as well (e. g. an homogeneous image collection of textures), by exploiting any a priori information that is available.

3. Furthermore, let us remind that the descriptors presented in Chapters 5 and 6 are all oriented towards a particular characteristic, such as colour or texture. Consider additionally the fact that these descriptors have varying degrees of pertinence for each retrievable concept, since for instance it would be rather irrelevant to use texture descriptors with sunset images, or anything other than texture with images of minerals. Thus, it is necessary to be able to determine dynamically the most adequate descriptor(s) for a given image category.

4. A further issue to take into consideration is the query form, as multiple choices are possible (Sec. 5.2.4). It would be to the user's interest to provide an intuitive method of query, as well as an easy to use interface.

5. And last, retrieval efficiency is another limitation, and a considerable one in this case since the proposed morphological approaches with their current implementations have been shown to be relatively time consuming.

In the light of these requirements, we will now present the proposed CBIR architecture.

7.3 Proposed architecture

Although an important portion of today's CBIR oriented efforts is based on unsupervised clustering of their visual databases [146], here we opt for a semi-supervised classification based approach using keyword associations, which will be shown to better suit the aforementioned design requirements. In particular, let us now elaborate on each of the design goals and restrictions, and note that their quantitative influence on the system are studied in Sec. 7.5.

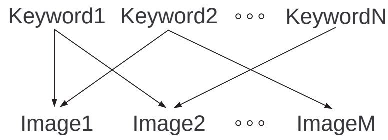


Figure 7.1: Keyword and image associations.

7.3.1 Semantic sensibility

In an ideal CBIR system, the semantic properties of the depicted objects are expected to be captured by the descriptors in use. However this great challenge is far from resolved, even though considerable advancement has been achieved through advanced descriptors, such as scale invariant interest points [176]. Since the morphology based descriptors presented in the previous chapter are rather focused on low-level image properties, and represent a direct application of colour morphological operators to content description, it would be deceivingly optimistic to expect them to provide at their state a semantically satisfying retrieval performance. Then again, nothing hinders the design of colour morphology based operators oriented towards interest point extraction, which constitutes a promising perspective in this regard. That is why, it has been decided to boost the process of semantic description through an external source, in other words by means of a learning set of example and counter example images, appropriately selected by an expert, and describing the possibly abstract concepts present in the database under consideration. Moreover, each retrievable concept is denoted by a textual keyword, hence leading to an initial association between keywords and groups of images (Fig. 7.1).

The use of keywords with (semi-)supervised annotation systems is of course no novelty, as it has been elaborated in Sec. 5.2.3. They have been used in combination with image regions [65], as well as with entire images [42] with a variety of classification methods [151, 148], thus they represent an approach which has relatively withstood the test of time and has been refined with various extensions exploiting the keyword-to-images associations [154]. In the present case, as we lack an effective and reliable segmentation method, we advocate the association of a given keyword with an image in its entirety. But of course the system is flexible enough to accommodate future modifications leading to the use of region based associations.

To further clarify the advantage of this approach in the light of the requirements mentioned in Sec. 7.2, consider a retrievable concept, described by an appropriate keyword, e. g. sunset. By providing the system with visual data effectively depicting sunsets as example images, one makes it possible for the system (e. g. by means of statistical approaches) to search for and find the common elements of these examples that define the concept under consideration. An even better accuracy in this last step can be achieved by providing additionally counter example images, for instance to avoid associating sunsets with just the presence of red-yellow since this may lead for example to the retrieval of red-yellow rose images. However, this step provides no guarantee that the system will effectively determine the elements distinguishing sunsets from all other possible concepts, and is largely dependent on the chosen example (EI) and counter example images (CEI). Nevertheless, with correctly chosen sets of EI and CEI, an expert becomes capable of describing to the system any concept whatsoever, hence bringing her own semantic knowledge into the system. Furthermore, even with badly chosen EI and CEI sets, the expert can still contribute and balance that step through relevance feedback, by judging the retrieval output generated by those EI and CEI sets, and modifying those sets accordingly.

Additionally, textual keywords may be organised into a hierarchy defined by the expert, hence allowing even more complicated inter-keyword semantic relations to be modelled. Besides, this potential of image description by keywords has been the main motivation behind the efforts oriented towards the development of complicated keyword ontologies [98, 100].

7.3.2 Adaptability

The notion of adaptability in this context denotes the capacity of the system to adapt itself to the heterogeneity of the database under consideration, so as to provide a satisfying annotation and retrieval performance. Naturally, it is a desired property in general, and in the present case of high priority as well.

CBIR systems based on unsupervised clustering of their image collections (auto-annotation), do not usually attempt to achieve this property directly, as the classification mechanisms that they possess (e. g. self-organising maps) provide an intrinsic adaptability capacity [136, 139], which however can lead to considerable performance variations, depending on the heterogeneity of the image collection under consideration. This variation is often attributed to the inability of the classifier to capture reliably, accurately and automatically higher level semantics [136]. That is why semi-supervised systems providing some sort of external "help" to the classifier have led to improvements in this regard [65]. More precisely, this "help" is usually in the form of example images [121], selected according to the heterogeneity level of the database under consideration, that are then used in order to teach the classifier how to differentiate the concept under study from others.

For example, consider a relatively heterogeneous database, where it would suffice to choose a few distinctive sunset images in order to differentiate them from accordions, cars, etc. Whereas with a more "difficult" database containing sunsets and various other content of red-yellow nature, it becomes necessary for the expert to use multiple counter example images, in order to avoid the confusion of sunsets with them, hence allowing the system to adapt its function to this database. Consequently, we opt here for the use of example and counter example image based learning.

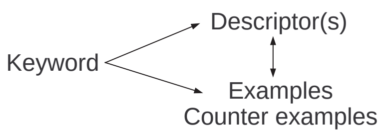


Figure 7.2: Figure showing the association of descriptors to keywords, EI and CEI sets to keywords, as well as the decriptor-example images relation, since the descriptor is chosen based on them.

7.3.3 Dynamic descriptor-keyword association

Furthermore, as elaborated in Sec. 5.2.3, each descriptor possesses different strengths and weaknesses, thus being appropriate for the description of certain concepts under certain conditions, while it might fail otherwise. Hence, various methods have been developed for choosing the most adequate descriptors for a given image collection or concept, depending on the architecture [66]. As a matter of fact, since its appearance within the MARS system [116, 233], *dynamic descriptor allocation* or *automatic feature selection* has come to be considered nowadays as a standard function of CBIR systems.

A keyword based annotation engine can easily accommodate this type of a function by associating each keyword, that represents in its turn a retrievable concept, with the descriptor(s) most suitable for its description; the suitability of which may be determined using the EI and CEI sets, provided by the expert (Fig. 7.2). In detail, given for instance examples and counter examples of sunsets, the system can use these learning sets for choosing the descriptor that leads to similar feature vectors for example images, and dissimilar feature vectors for counter example images. Hence, by using feature vectors optimised for the described concept, it contributes to the overall description and adaptability capacity of the system.

7.3.4 Ease of use

An additional advantage of a textual keyword based approach is the fact that it offers to users an already familiar retrieval means, to which they are accustomed to from text-based search engines. Typing the keyword "sunset" to search for sunset images is obviously easier than having to look first for a sunset image that will be used for querying by example. The drawbacks of alternative query methods have been elaborated in Sec. 5.2.4.

7.3.5 Efficiency

Besides, since the associations among keywords and image groups are formed offline, under the supervision of the expert, the users can then exploit these static associations online, hence enjoying constant time retrieval performance. Consequently the burden of processing and memory complexity of morphological operators, is shifted entirely to the stage of administration by the expert, making it transparent for the user, whose available options include a read-only exploitation of these associations.

7.3.6 Drawbacks

Even though the aforementioned design choices have various advantages to offer, they also suffer from certain drawbacks. First and foremost is the need for an expert who will have to define each keyword to be used by the system. And of course this assumes that the expert already knows the concepts present within the image collection the system will be working with. This is obviously a serious limitation, but one that we had to impose in order to obtain an engine able of a reliable annotation capacity, as it was demanded by the project sponsors. May it be a collection of carpet textures, minerals or museum artwork, being able to accommodate the expert's knowledge of the "field" becomes invaluable in this regard.

Unlike unsupervised clustering schemes, where the system attempts to determine entirely on its own semantically uniform image groups, here it is a process relying partially on an external intervention. While this makes it possible for the expert to direct the system's behaviour according to her specific needs along with the image collection under consideration, thus resulting in a more adaptable system, it also unavoidably brings in her subjective view for each keyword's definition, which may very well be different from that of users. Of course, this situation may be attenuated with the presence of multiple experts.

Furthermore, the definition process of a keyword is also highly dependent on the particular EI and CEI sets that have been chosen (Sec. 7.5). For instance in order to define the keyword “truck” in a database containing multiple vehicle types, it is imperative to use counter example images that will differentiate trucks from cars and buses. In other words, the expert’s expertise is of crucial importance, hence making the system only as performant as the expert administering it, since it is not designed to be an independent annotation and retrieval engine, but rather a tool that will aid the expert exploit her knowledge of the image domain under consideration while avoiding manual annotation. Moreover, as the descriptors in use have a direct influence on the overall performance of the system, in terms of capturing the semantic nuances of the provided examples, the morphological descriptors proposed in the previous chapter represent a drawback in this regard due to their global character, as well as a future perspective, consisting of developing localised content descriptors in combination with a reliable and automatic segmentation method.

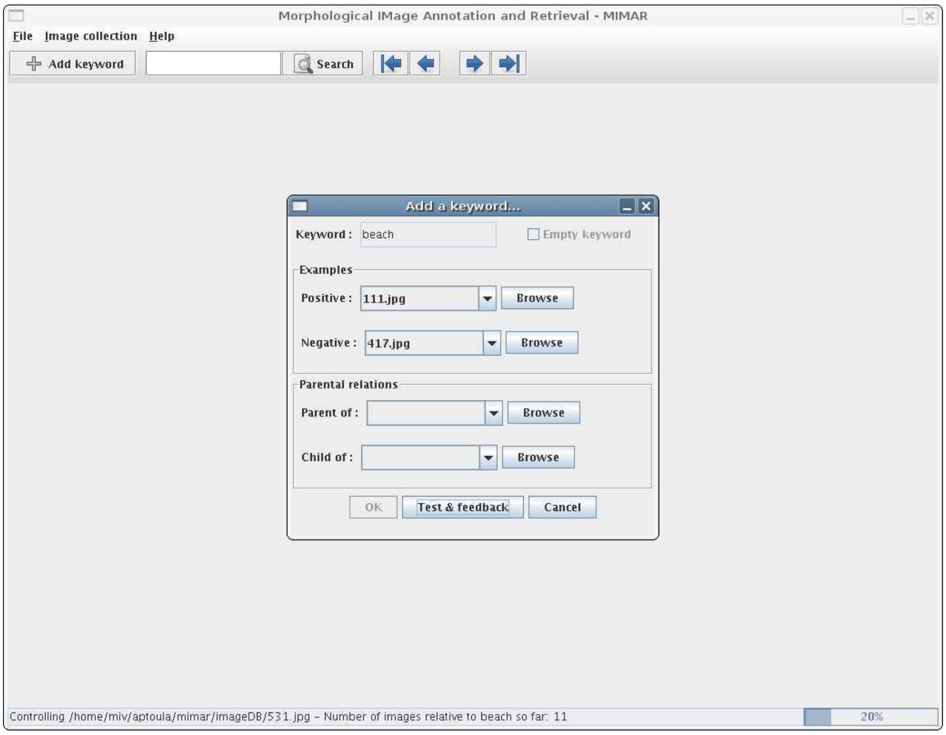


Figure 7.3: Screenshot of the MIMAR keyword adding interface.

7.4 Implementation

In this section we present the software that has been implemented based on the design principles presented in Sec. 7.3. It should be noted that the software has been developed at this late stage of my thesis with the intention to obtain a working prototype of the previously mentioned architecture, where morphological descriptors may be tested and evaluated. That is why functions have been kept at the bare minimum, and the main focus has been on obtaining a flexible and extensible structure, that can be easily modified in the near future in order to provide the project sponsors with a fully operational CBIR system, incorporating most if not all of state-of-the-art elements. In particular, our architecture has been implemented within a Java-based framework, in order to produce a piece of software named **M**orphological **IM**age **A**nnotation and **R**etrieval (MIMAR). MIMAR may be used in two modes, expert and user, and while the first has full access to all functions, the second is limited with only querying by keyword(s). Let us now detail each of the supported operations.

7.4.1 Adding a keyword

Initially the software is assumed to start working with an unannotated image collection. The form of this collection, either in the file system, or as a database is of no importance in this context, as we concentrate on the inner workings of the software and not on its interaction formalities with the actual data. For precision's sake, note that the current implementation of MIMAR functions with image collections present within a standard directory of the file system. Given this set of images, the expert can initiate the only possible function, which consists in adding a new keyword (Fig. 7.3).

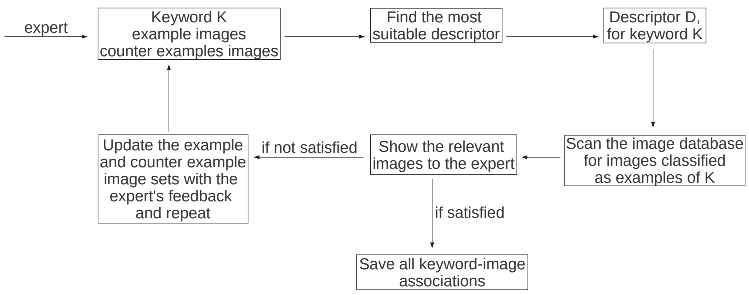


Figure 7.4: Diagram of steps for adding a new keyword into MIMAR.

The diagram of this procedure is shown in Fig. 7.4. In particular, in order to define a keyword K, the expert must provide a set of EI along with CEI, that represent K according to her judgement. There is no upper limit to the number of images, however at least 2 of each are required. Optionally, in order to form a hierarchy of keywords, and hence gradually form an ontology specific to the image collection under consideration, the "children" and/or "parents" of K may also be set at this stage. The parenthood relations among keywords represent intuitively concept inclusions. More precisely, children and parents of K are respectively keywords that denote subsets and supersets of K. For instance the children of "flower" could be "rose" and "daffodil", while "plant" could be one of its parents. From a computational point of view, these relations describe a connected graph structure as for instance the one illustrated in Fig. 7.5.

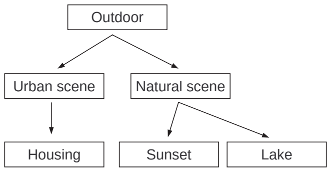


Figure 7.5: Example of keyword hierarchy.

Next, we determine the best suited descriptor D among the available options, which include all the proposed as well as tested descriptors of Chapter 6. Of course, the system itself is independent from the descriptors set, and can function with any feature extraction means equipped with a similarity metric. To explain, we attempt to determine the descriptor that "best describes" the keyword K based on the given images, where the quality of description is denoted by its capacity to group the feature vectors of example images in the multi-dimensional space, while separating them from those of counter example images. Formally, following a distance normalisation to $[0, 1]$, we minimise the quotient of intra-example distances divided by inter-example distances:

$$\mathrm{c} = \frac{\text{intra-example distance} \times \text{number of examples}}{\text{inter-example distance} \times \text{number of counter examples}} \tag{7.4.1}$$

This formulation also takes into account the sizes of EI and CEI sets. Thus, the eventual damaging effect of EI and CEI sets with very different cardinalities can be partially balanced. More precisely, in case for instance the number of example images is very low, then it is mainly their distance to counter examples that determines the best descriptor. Of course looking for a single descriptor with fixed arguments is clearly not the best possible approach, since the concept under consideration might possibly require information of various types (e. g. interest points, texture, colour, etc) in order to be effectively distinguished. In which case this step would be modified as a search not for the best descriptor, but rather for the optimised weight of each descriptor along with its arguments, and the actual optimisation can be realised by any multi-criterion optimisation approach, e. g. genetic algorithms. Due to time constraints this extension has been left as a priority perspective.

Having determined the best way to calculate features, then we scan the image collection, and compute the feature vectors of all images, using the chosen descriptor D. Naturally, this step is computationally most expensive, depending on the size of the collection under consideration as well as on D. For instance entering a keyword of the Caltech-256 dataset with 10 examples and 10 counter examples using an average laptop can require up to several days depending on the chosen descriptor's complexity (e. g. up to 6 days with the texture descriptor of Eq. (6.3.4)). For each examined image I, we need to decide if it will be associated or not with K. This is a discrete decision, an image I is deemed either relevant at some fuzzy degree to K, in which case it is associated with it, or irrelevant and is discarded. This decision may be realised with any classification method e. g. k nearest neighbours (kNN), support vector machines (SVM), etc. Using such a classifier C, we associate an image I with the keyword K, if its feature vector is classified or labeled as belonging to the feature vectors of the set of EI. In which case the *relevance* $\in [0, 1]$ of I with respect to K is set as:

$$\text{relevance} = 1.0 - \text{distance of I's feature vector to that of example images} \tag{7.4.2}$$

Otherwise (i. e. I belongs to CEI), I is considered as irrelevant to K and ignored. Following the calculation of relevances, the images relevant to K are shown to the expert in sorted order who can then judge the results. If satisfied, the process is ended by saving all associations of images to K, and updating the parent keywords, if any, of K to include the newly associated images. Consequently, the keyword structure is formed by the following information:

$$\text{Keyword} = \{ \begin{array}{l} \text{keyword}, \\ \text{descriptor}, \\ \text{children}, \\ \text{parents}, \\ \text{feature vectors of example images}, \\ \text{feature vectors of counter example images} \; \} \end{array} \tag{7.4.3}$$

If not satisfied, then the expert may indicate positive and negative results, hence providing her feedback (Fig. 7.6). This feedback is exploited by updating the sets of EI and CEI to include respectively the indicated positive and negative results, hence improving the classification accuracy. Then, the entire process is repeated by recomputing the most suitable descriptor for the updated sets of example and counter example images.

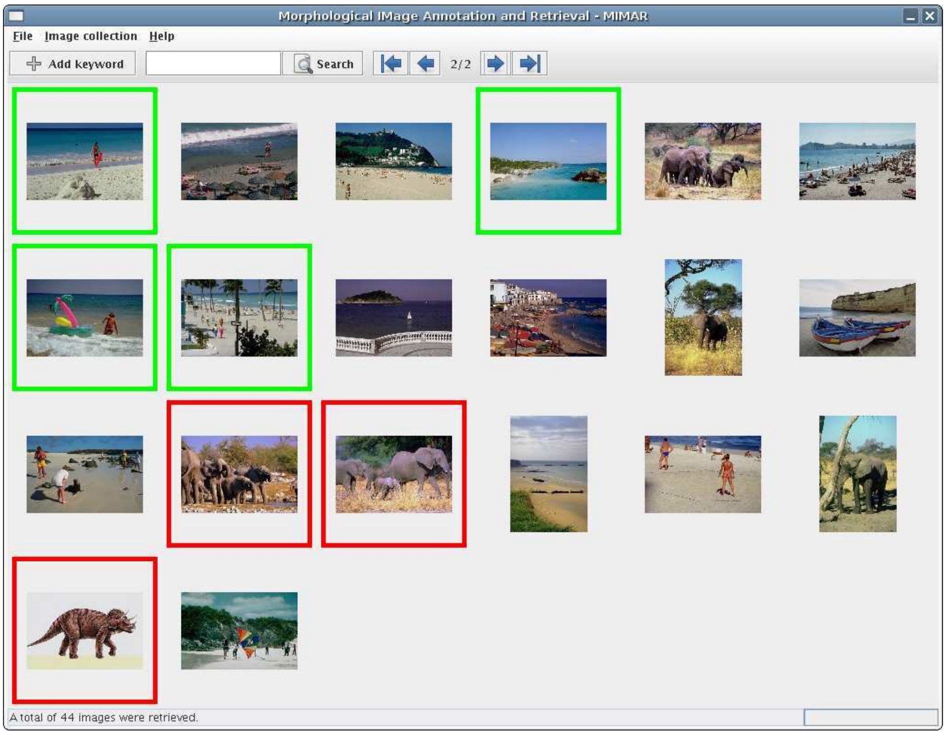


Figure 7.6: Screenshot of the MIMAR relevance feedback interface.

7.4.2 Updating the image collection

Image collections are volatile entities, where given today's digital storage capacities, adding new images is a an operation realised more often than their deletion. In case the collection is modified, either because some image has been added or eliminated, it becomes necessary to update the keyword-images association network. By launching this process, any images associated with keywords that are not found physically in the collection are eliminated from their respective keywords, as well as from the parents of these keywords recursively. As to newly added images that have not yet been associated with any keyword, each of them is tested against all available keywords in the system, in an attempt to determine if they fit the description of that keyword. To realise this step, the feature vector of an image I under consideration is computed using the descriptor D associated with the keyword K against which it is compared. Then if its feature vector is classified as belonging to the set of EI of K, it is associated with K according to Eq. (7.4.2). Consequently, it is only necessary to compute the feature vectors of the new images, with no requirement to recalculate for the entire association network.

7.4.3 Retrieving a keyword

Keyword retrieval is the only function offered to non-expert users. Considering that the entire system design is oriented with the goal of providing an efficient, effective and intuitive retrieval, this operation is trivially simple for the user, since it suffices to enter a keyword present in the system (Fig. 7.7). Then with constant time complexity (less than 1 sec in practice), the retrieval is realised by showing the images associated with the sought keyword. Moreover, through the relations of parenthood among keywords, the system design makes it trivial to realise in the future an extension supporting queries through keyword combinations based on simple boolean operations such as "house and flower" (i. e. retrieve images associated with both "house" and "flower") or "flower but not rose" (i. e. retrieve images associated with "flower" but not with "rose").

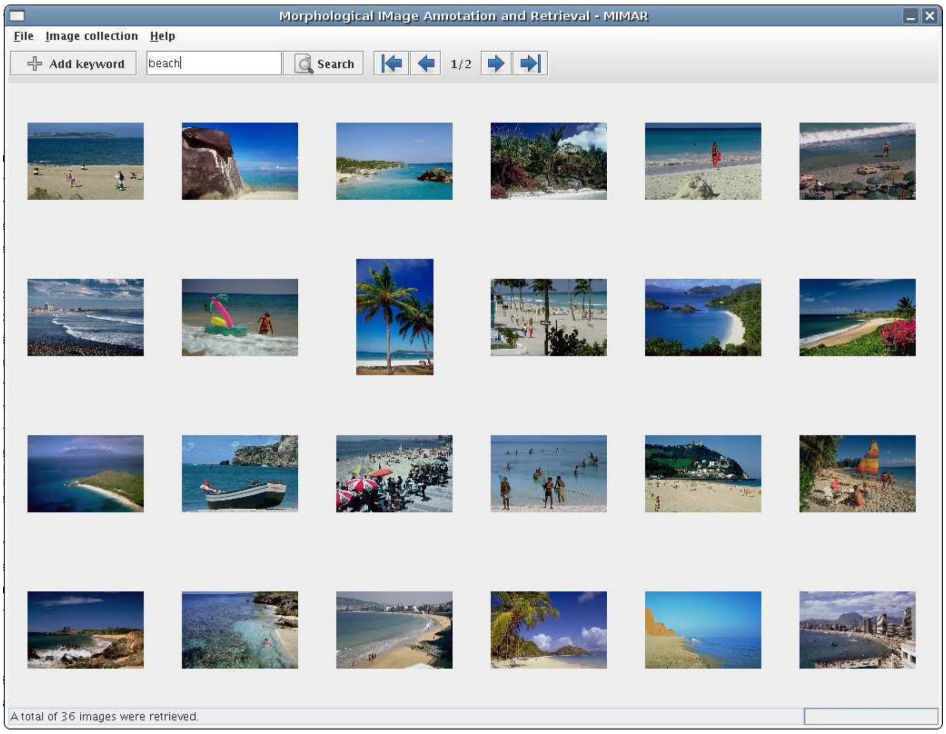


Figure 7.7: Screenshot of the MIMAR interface after retrieval.

7.5 Experimental evaluation

This section presents the results of a series of retrieval tests, that have been carried out in order to illustrate the principal functions of the implemented system, and assert the remarks that have been made previously in this chapter with experiments. To this end, we have chosen the Caltech-256 image collection [88], a recent and most challenging image database, with 256 object categories and at least 80 images per category, making a total of approximately 30.000 images. MIMAR has been set to use the kNN classifier for annotation, with $k = 1$ and the Euclidean distance. kNN is an elementary option in this regard, unlike SVM, hence better illustrating the remaining elements of the system. Moreover, all descriptors have been set to the configurations used in Secs. 6.2.4 and 6.3.2. We start by studying the impact that EI-CEI sets have on the system's performance.

7.5.1 Influence of example sets on annotation

This first test aims to control the level at which the chosen example and counter example image sets influence the overall keyword to image association process. In other words, considering that the EI-CEI sets are defined by the expert, it aims to measure the effect of the expert's choices on the system's performance. We break up this problem into two parts:

1. given a keyword K, what is the effect of badly chosen EI-CEI sets,
2. and what is the effect of the cardinality of the EI-CEI sets?

To this end, we have first selected 10 object categories: elephant, crab, leopards, laptop, starfish, motorbike, brain, sunflower, bonsai and grand-piano. As far as the first part of this test is concerned, we have tried annotating repeatedly the image collection for each of these 10 keywords, using 4 images with 2 examples picked randomly among the images of the same category and 2 counter-examples that have been also picked randomly at each time from the rest of the image collection, in order to observe the quality of annotation depending on the learning set. The quality of annotation has been measured using the category specific annotation accuracy (i. e. percentage of correctly annotated images among the retrieved ones), and of course no relevance feedback has been provided, so as to isolate the effect of the learning process. Table 7.1 shows the average annotation accuracy that has been obtained after 100 annotations as well as the standard deviations.

Elephant	Crab	Leopards	Laptop	Starfish
23.6 ± 3.2	17.6 ± 3.5	12.2 ± 4.0	15.5 ± 4.3	15.2 ± 4.15
Motorbike	Brain	Sunflower	Bonsai	Grand-piano
11.5 ± 2.1	18.8 ± 3.25	16.4 ± 2.75	17.7 ± 2.35	14.3 ± 3.1

Table 7.1: Average annotation accuracies and their deviations for 10 image categories of Caltech-256, obtained over 100 annotations with random examples (2) and counter examples (2).

According to the obtained values, one can observe relatively high deviations from the average annotation accuracy of each class. These high values indicate the importance of the chosen images for defining a given keyword, and indirectly emphasise the responsibility of the expert in this regard. Moving to the next sub-question, we have used the same 10 keywords to annotate the image collection, but this time using EI-CEI sets of various sizes. More precisely, for each size $|EI| + |CEI| = k$, $k \in \{4, 35\}$, we have tried 100 randomly formed and distinct EI-CEI sets, and computed the accuracy of the first retrieval results, once more with no feedback whatsoever. Fig. 7.8 shows the obtained results, where one can clearly observe the positive effect of using larger example and counter example sets. Based on this result we can also presume with certainty that relevance feedback, either in the form of indicating more examples or counter examples, will result in the same type of improvement in terms of performance. Because, feedback within MIMAR is realised by updating the EI-CEI sets (Sec. 7.4.1), hence it is identical in effect to using EI-CEI sets of increased sizes as in this particular case.

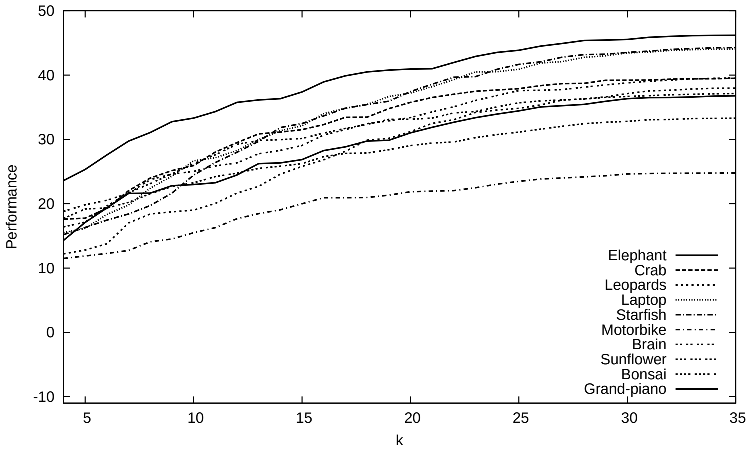


Figure 7.8: Annotation accuracies over 100 annotations with random EI-CEI sets of various sizes $k = |EI| + |CEI|$.

7.5.2 How much does dynamic keyword-descriptor association help?

Using a descriptor suitable to the processed images is certainly a positive step as far as their effective characterisation is concerned [233]. However, how much does this step contribute, and furthermore how effective is the criterion in Eq. (7.4.1) that we have chosen to use ? To answer these questions, we have employed once more the 10 keywords of Caltech-256 (Sec. 7.5.1), that have been added to the system repeatedly, using all the available descriptors with randomly picked 2 example as well as 2 counter example images. Table 7.2 shows the annotation accuracies that have been obtained over 100 repetitions for each descriptor-keyword couple, note that once more there has been no feedback, as these values are based on the first retrieval results.

Judging from the obtained values, it is clear that the dynamically associated descriptor is the best possible among the available choices for all of the tested keywords. This result not only constitutes an indication that the criterion used for choosing the most suitable descriptor is practically useful, despite its simplicity, but mainly it shows that dynamic descriptors possess indeed a practical interest not to be ignored. One would naturally expect the benefit to be even higher once keyword specific descriptor arguments are used (e. g. adaptively set scale number to MHL, or SE sizes for granulometries).

	Histogram	AC	CSD	MHW	CSG	MHL	Gra	Cov	Combo
Elephant	12.9	19.8	23.2	21.7	22.8	***23.9***	16.2	17.6	20.6
Crab	20.8	22.9	31.8	34.9	35.3	***36.2***	21.4	23.5	23.6
Leopards	3.3	6.2	8.8	10.2	***12.2***	10.8	3.3	4.7	5.4
Laptop	12.8	17.9	***26.1***	25.5	24.7	24.6	13.1	15.1	16.1
Starfish	23.9	28.8	30.5	***35.9***	34.7	33.2	26.1	23.6	23.1
Motorbike	18.6	20.8	23.4	***31.6***	30.9	30.3	21.1	20.8	22.2
Brain	10.7	11.6	13.6	25.5	24.4	***28.5***	17.6	18.1	18.8
Sunflower	20.6	22.8	25.1	32.4	33.3	***36.6***	23.2	25.5	28.8
Bonsai	27.7	30.9	32.6	34.1	***37.2***	36.6	16.8	18.9	20.2
Grand-piano	11.8	13.6	20.9	31.4	33.7	***24.9***	12.8	14.7	15.9

Table 7.2: Average annotation accuracies over 20 annotations using all the available descriptors for each of the 10 keywords of Caltech-256. The best couples are in bold and the couples corresponding to the dynamically associated descriptors are in italic.

7.5.3 Is MIMAR adaptive?

The concept of adaptability has been detailed in Sec. 7.3.2, and although it is an important notion in the context of CBIR, it is rather hard to formalise, and as such it is also difficult to measure. Returning to the definition, the notion of adaptability in this context denotes a system's capacity to adapt itself to the heterogeneity of the image collection under consideration, in other words, an adaptive system should be able to function (i. e. annotate and retrieve) at a satisfying level, independently from the collection's heterogeneity as far as its content is concerned. In which case, one needs to ask: is MIMAR adaptive as demanded by the project goals (Sec. 7.2)?

To answer this question, we have taken two image collections of different heterogeneity in terms of content, Caltech-256 as heterogeneous and Outex13 (Sec. 3.4.2) as homogeneous. A non adaptive annotation system, would have problems with these two, since perceptually speaking the texture classes of Outex13 (Fig. 3.7) are clearly more similar to each other than the objects of Caltech-256 (e. g. elephant, crab, laptop, etc) are. Consequently, a non adaptive system with a static annotation capacity would function sufficiently well with at most one of them, being unable to either separate the texture classes of Outex13 or recognise as similar the various objects of Caltech-256.

	MIMAR (no feedback)	MIMAR (1 round: +1 EI, +1 CEI)	Best
Caltech-256	29.3 ± 1.6	30.1 ± 1.7	59.84 Varma [274]
Caltech-256 (Combo)	11.1 ± 1.3	12.3 ± 1.4	-
Outex13	74.3 ± 2.5	82.1 ± 2.6	94.7 Ojala et al. [191]
Outex13 (MHL)	31.5 ± 3.2	33.3 ± 3.2	-

Table 7.3: Classification results of MIMAR for Caltech-256 and Outex13.

In particular, we have carried out 20 annotations for each keyword of Caltech-256, using randomly selected 15 examples and 10 counter examples for each image category at each annotation. This has been realised once using the dynamic keyword-descriptor associations and once with the combined (Combo) texture descriptor of Eq. (6.3.4) (same configuration as in Sec. 6.3.2) that has provided the best results in the tests of the previous chapter. Our motivation for this additional test has been to observe the effect of relatively unsuitable descriptors on the overall image collection. Moreover the classification result of Varma [274] has been obtained using 20 training examples. For Outex13, we have carried out the same tests, using 5 examples and 3 counter examples, since each texture category contains only 20 images. Likewise, we have repeated this experiment for the MHL operator (same configuration as in Sec. 6.2.4) that has provided relatively satisfying results with non textural image data in the previous chapter. The result reported by Ojala et al. in Ref. [191] has been obtained using 10 training examples with an identically configured kNN classifier. Table 7.3 shows the average classification performances over 20 annotations of all image categories, obtained for both image collections, as well as the best reported values for them. The values obtained with no feedback are of importance, since they are comparable to the best known results, obtained by semi-supervised clustering based systems.

In general one can observe that MIMAR leads to clearly inferior results with respect to the state-of-the-art. Nevertheless, considering that our system's main strength is its novel descriptor set, whereas almost all other design elements are in their basic form (e. g. classifier, descriptor choice), one cannot truly say that the difference is that great. Furthermore, MIMAR has been designed to be an aid for annotation to the expert, and not to undertake the entire task, hence its full potential cannot be achieved by a single annotation run, which we can also observe by the positive effect of a single feedback with +1 EI and CEI. What is most important though, is that these results provide an indication that MIMAR can indeed adapt to these image collections of relatively different content, thus partially satisfying the project goal of adaptability.

7.6 Conclusion

The objective of this chapter has been to combine the findings of the second part of this thesis in terms of morphological content descriptors, in order to provide a CBIR system with a morphological foundation illustrating the practical potential of these operators. To this end, we have developed an architecture based on a number of goals and limitations, some of which were imposed by the financial sponsors of this project. Specifically, we have aimed to obtain a system capable of queries at a semantic level, adaptable to different image domains and capable of exploiting any a priori information. Additionally, given the computational complexity of the proposed descriptors, it became necessary to find a way to provide not only effective but an efficient retrieval means as well.

With these goals, we have oriented our efforts towards a keyword-based architecture, named MIMAR, where queries are realised by means of textual keywords, thus providing an intuitive retrieval interface. The keyword-image associations are realised through learning from example and counter example image sets, that are provided to the system by an expert, who thus brings into the system her semantic knowledge related to the image domain under consideration. Furthermore, by using dynamic descriptor choices for any given keyword, we manage to optimise the keyword-descriptor associations, hence avoiding the pitfall of using descriptors that are inadequate for a given image category. On the other hand, the necessity for an expert is also the main weakness of this design, since her knowledge, all be it invaluable, remains subjective and may differ from the views of the users for a given concept. Moreover, the system's capacity to determine the images within the database relative to a new keyword, is highly dependant on the quality of the provided examples and counter examples, in other words on the expert's expertise.

All the same, MIMAR is open to a number of improvements that hold the potential of significantly improving its performance. Among others, one can mention its flexibility to function with any adequate classifier, support for using combinations of optimal descriptors as well as descriptor arguments, the use of localised/region based descriptors and finally the implementation of advanced queries by means of boolean operator based keyword combinations.

Chapter 8

General conclusion

Throughout this thesis, various approaches have been presented, dealing primarily with the extension of morphological operators to colour images, as well as with their application to content-based colour image retrieval. This chapter provides an overview of the contributions made, along with some future perspectives.

8.1 Contributions of this thesis

Concerning the first part of this thesis, where the problems of colour morphological analysis have been elaborated, we have initiated our study with a coupled presentation of colour space options and of the state-of-the-art in colour, and extensible-to-colour, multivariate morphological approaches. A comparative study of which has been provided using noise reduction and texture classification tests, that led us to realise the importance of the notion of flexibility from the part of the ordering approach, since the vast variety of practical needs, render finding a single colour ordering appropriate for all tasks considerably difficult. The same results have made it possible to justify our choice of using the lexicographical ordering in combination with phenomenal spaces, and in particular with LSH.

Next, we concentrated on the specific difficulties of defining colour morphological operators through lexicographical ordering in a phenomenal colour space, namely the ordering of the hue component and controlling the extreme prioritisation effect of lexicographical comparison cascades. In particular, it was judged that distances with respect to a single hue reference are not practically sufficient in cases where multiple dominant hues are present within images. Consequently, it has been proposed to use a multitude of hue references in a similar way, obtained automatically. Furthermore, based on the principle of flexibility, we have proposed three variations to lexicographical ordering with user defined arguments, in an attempt to attenuate and control its prioritisation effect. Specifically, a generalisation of the α-modulus lexicographical ordering has been presented, that can accommodate arbitrary priority distributions by means of spectral sub-quantisations, hence making it possible to model the dependency relations among brightness, saturation and hue. An additional variation includes the use of morphologically flattened images as markers, in order to determine the image regions where lexicographical cascades are supposed to shift to latter dimensions. Thus, it is shown that the tool of sub-quantisation can be used also in a spatial context with the same purpose. The third variation consists of α-trimmed lexicographical extrema, a means of computing multivariate extrema from a set of vectors, which offers a way of controlling the contribution of each dimension to the lexicographical extrema computation process. Nevertheless, as there is no underlying ordering it results in pseudo-morphological operators. All three approaches have then been tested with colour image applications where competitive and often superior performances have been obtained with respect to state-of-the-art orderings. Moreover, as an example of the findings of this part, the hit-or-miss transform has been studied, and defined for colour and more generally for multivariate images.

In the second part, having singled out the generalised spectral sub-quantisation based lexicographical ordering model in the LSH colour space for colour morphological operators, we proceeded to study the field of content-based image retrieval. By means of the overview of this domain, the main challenge of bridging the semantic gap has been emphasised. Then we have undertaken the task of developing morphological content descriptors, of which three colour and one texture oriented have been introduced. In particular, following the identification of fundamental colour properties, in the context of CBIR, we have focused our efforts to capturing the spectral as well spatial colour distribution of images. To explain, a colour specific granulometry has been first studied, aiming to provide an independent size distribution for each colour. Then two morphological approaches have been tested for forming multi-scale histograms, namely morphological levelings and the watershed transform. As to texture, we have aimed to obtain an improved combination of the two principle morphological texture descriptors: granulometry and morphological covariance. To this end, both operators have been fused through the use of multiple SE variables. All proposed descriptors have been tested using precision-recall charts and ANMRR measurements, showing competitive performances, which however are obtained at a relatively high computational cost.

The aforementioned descriptors have then been implemented within a CBIR system named MIMAR, which has been designed according to the goals and limitations set by our project sponsors, aiming to provide support for semantic retrieval, while also being able to adapt its performance to domain specific image databases, and functioning with generic cases. With this purpose, we opted for a keyword-based architecture with a supervised learning mechanism, using the example and counter example images provided by an expert. Standard design elements such as relevance feedback and dynamic keyword-descriptor association mechanisms have also been included. Thanks to the amount of semantic knowledge brought to the system by the expert in the form of learning sets, as well as by means of her feedback, the project goals have been partially achieved. Nevertheless, the subjectivity as well as level of expertise of the expert remain the system's main weakness.

8.2 Perspectives

As far as the first and mainly theoretical part of this thesis is concerned, the proposed ordering methodologies have been designed to be primarily flexible, so as to be able to adapt to multiple practical uses. As this has been achieved partially by means of user defined arguments, it has resulted in the new challenge of setting these arguments in a robust and automatic way, if possible, in order to achieve an even higher level of adaptability. Moreover, all three lexicographical ordering variations have been designed independently from the context of colour images, to which they have been later applied. Hence they hold the potential of providing a means for vector multi-spectral morphological image analysis, such as for remote sensing and astronomical data, where lately the morphological approaches have been gaining momentum. Besides, a colour hit-or-miss transform can also be of considerable practical use, as it is in the binary and grey-scale case, once its robustness to scale variations and view point changes is sufficiently developed.

From the point of view of content-based image retrieval, a multitude of future perspectives can be observed, such as the further development of the proposed descriptors, in particular in terms of execution speed. Additionally, we have failed to study the family of shape descriptors, due to both time constraints as well as to the lack of a reliable and automatic colour segmentation method. Furthermore, once a semantically valid partitioning of the input is obtained, one may proceed to the development of localised descriptors, that focus on individual objects rather than the totality of their input. Similarly, a morphological shape descriptor compatible with colour images, able to exploit brightness, saturation and hue related borders can be considered; while another particularly interesting future research corridor consists of the implementation of more advanced content descriptors, such as interest points based on colour morphology approaches. As to MIMAR, considering its basic present form, it can further incorporate a variety of modifications, including but not limited to support for using combinations of optimal descriptors as well as descriptor arguments, the use of localised/region based descriptors and finally the implementation of advanced queries by means of boolean operator based keywords combinations.

List of Publications

Articles in refereed international journals

1. Erchan Aptoula and Sébastien Lefèvre. α-Trimmed lexicographical extrema for pseudo-morphological image analysis. *Journal of Visual Communication and Image Representation*, 19 (3):165-174, April 2008.
2. Erchan Aptoula and Sébastien Lefèvre. On lexicographical ordering in multivariate mathematical morphology. *Pattern Recognition Letters*, 29 (2):109-118, January 2008.
3. Erchan Aptoula and Sébastien Lefèvre. A comparative study on multivariate mathematical morphology. *Pattern Recognition*, 40 (11):2914-2929, November 2007.

Articles in refereed international conference proceedings

1. Erchan Aptoula and Sébastien Lefèvre. A basin morphology approach to colour image segmentation by region merging. *Proceedings of the Asian Conference on Computer Vision (ACCV'07)*, volume 1, pages 935-944, Tokyo, Japon, November 2007.
2. Erchan Aptoula and Sébastien Lefèvre. On morphological colour texture characterization. *Proceedings of the International Symposium on Mathematical Morphology (ISMM'07)*, pages 153-164, Rio de Janeiro, Brazil, October 2007.
3. Erchan Aptoula and Sébastien Lefèvre. Pseudo multivariate morphological operators based on α-trimmed lexicographical extrema. *Proceedings of the IEEE International Symposium on Image and Signal Processing and Analysis (ISPA'07)*, Istanbul, Turkey, September 2007.
4. Erchan Aptoula and Sébastien Lefèvre. Spatial morphological covariance applied to texture classification. *Proceedings of the International Workshop on Multimedia Content Representation, Classification and Security (MCRCS'06)*, Lecture Notes in Computer Science, volume 4105, pages 522-529, Istanbul, Turkey, September 2006.
5. Erchan Aptoula, Sébastien Lefèvre and Christophe Collet. Mathematical morphology applied to the segmentation and classification of galaxies in multispectral images. *Proceedings of the European Signal Processing Conference (EUSIPCO'06)*, Florence, Italy, September 2006.

Articles in refereed national conference proceedings

1. Erchan Aptoula and Sébastien Lefèvre. Pseudo-opérateurs morphologiques multivariés basés sur les extrema lexicographiques α-tronqués. *ORASIS'07*, Obernai, France, June 2007.

List of Submissions and Reports

Articles in refereed international journals

1. Erchan Aptoula and Sébastien Lefèvre. On the morphological processing of hue. Image and Vision Computing (submitted), 2007.

Technical Reports

1. Erchan Aptoula, Sébastien Lefèvre and Christian Ronse. A hit-or-miss transform for multivariate images. *Technical report.* March 2008.

2. Erchan Aptoula and Sébastien Lefèvre. Morphological descriptors for content-based image retrieval. *Technical report.* December 2007.

3. Erchan Aptoula. Image collections for content-based image retrieval. *Technical report.* July 2007.

4. Erchan Aptoula. Content-based image retrieval - an overview. *Technical report.* July 2006.

List of Algorithms

List of Figures

List of Tables

Bibliography

[1] H. M. Al-Otum. Morphological operators for color image processing based on Mahalanobis distance measure. *Society of Photo-Optical Instrumentation Engineers*, 42(9):2595–2606, September 2003.

[2] H. M. Al-Otum and M. T. Ureikat. Color image morphology using an adaptive saturation-based technique. *Optical Engineering*, 43(6):1280–1292, June 2004.

[3] C. N. Anagnostopoulos, A. Koutsonas, I. Anagnostopoulos, V. Loumos, and E. Kayafas. Tile classification using the CIELAB color model. In *Proceedings of the International Conference on Computational Science*, pages 695–702, Atlanta, USA, May 2005.

[4] F. A. Andaló, P. A. V. Miranda, R. D. S. Torres, and A. X. Falcão. A new shape descriptor based on tensor scale. In U. M. Braga-Neto G. J. F. Bannon, J. Barrera, editor, *Proceedings of the 8th International Symposium on Mathematical Morphology*, volume 1, pages 141–152, Rio de Janeiro, Brazil, October 2007. INPE.

[5] I. Andreadis, M. Vardavoulia, G. Louverdis, and N. Papamarkos. Colour image skeletonisation. In *Proceedings of the 10th European Signal Processing Conference*, volume 4, pages 2389–2392, Tampere, Finland, September 2000.

[6] J. Angulo. *Morphologie Mathématique et Indexation d'Images Couleur. Application à la Microscopie en Biomédicine.* PhD thesis, Ecole des Mines, Paris, France, 2003.

[7] J. Angulo. Morphological color processing based on distances. Application to color denoising and enhancement by centre and contrast operators. In *Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'2005)*, pages 314–319, Benidorm, Spain, September 2005.

[8] J. Angulo. Unified morphological color processing framework in a lum/sat/hue representation. In C. Ronse, L. Najman, and E. Decencière, editors, *Mathematical Morphology: 40 Years On*, volume 30 of *Computational Imaging and Vision*, pages 387–396. Springer-Verlag, Dordrecht, 2005.

[9] J. Angulo. Morphological colour operators in totally ordered lattices based on distances: Application to image filtering, enhancement and analysis. *Computer Vision and Image Understanding*, 107(1-2):56–73, July 2007.

[10] J. Angulo and J. Serra. Morphological color size distributions for image classification and retrieval. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 46–53, Ghent, Belgium, September 2002.

[11] J. Angulo and J. Serra. Morphological coding of color images by vector connected filters. In *IEEE Proceedings of the 7th International Symposium on Signal Processing and Its Applications (ISSPA'2003)*, volume 1, pages 69–72, Paris, France, July 2003.

[12] J. Angulo and J. Serra. Modelling and segmentation of colour images in polar representations. *Image and Vision Computing*, 25(4):475–495, April 2007.

[13] E. Aptoula and S. Lefèvre. A basin morphology approach to colour image segmentation by region merging. In *Proceedings of the Asian Conference in Computer Vision*, volume 4843 of *Lecture notes in computer science*, pages 935–944, Tokyo, Japan, November 2007.

[14] E. Aptoula, S. Lefèvre, and C. Collet. Mathematical morphology applied to the segmentation and classification of galaxies in multispectral images. In *Proceedings of the 14th European Signal Processing Conference*, Florence, Italy, 2006.

[15] L. H. Armitage and P. G. B. Enser. Analysis of user needs in image archives. *Journal of Information Science*, 23(4):287–299, 1997.

[16] Y. A. Aslandogan, C. Thier, C. T. Yu, J. Zou, and N. Rishe. Using semantic contents and wordnet in image retrieval. In *Proceedings of the 20th international ACM SIGIR conference on Research and development in information retrieval*, pages 286–295, Philadelphia, USA, September 1997.

[17] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *IEEE Proceedings*, 78(4):678–689, April 1990.

[18] A. Aubert, D. Jeulin, and R. Hashimoto. Surface texture classification from morphological transformations. In J. Goutsias, L. Vincent, and D. S. Bloomberg, editors, *Proceedings of the International Symposium on Mathematical Morphology*, volume 18 of *Computational Imaging and Vision*, pages 253–252, Palo Alto, USA, 2000.

[19] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Jain, and C. F. Shu. The Virage image search engine: An open framework for image management. In *Proceedings of SPIE, Storage and Retrieval for Still Image and Video Databases IV*, pages 76–87, San Jose, USA, February 1996.

[20] C. Barat, C. Ducottet, and M. Jourlin. Pattern matching using morphological probing. In *Proceedings of the 10th International Conference on Image Processing*, volume 1, pages 369–372, Barcelona, Spain, 2003.

[21] V. Barnett. The ordering of multivariate data. *Journal of the Statistical Society A*, 139(3):318–355, 1976.

[22] I. Bartolini, P. Ciaccia, and M. Patella. Warp: Accurate retrieval of shapes using phase of Fourier descriptors and time warping distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.

[23] S. Batman and E. R. Dougherty. Size distributions for multivariate morphological granulometries: texture classification and statistical properties. *Optical Engineering*, 36(5):1518–1529, May 1997.

[24] P. Bhattacharya, W. Zhu, and K. Qian. Shape recognition method using morphological hit-or-miss transform. *Optical Engineering*, 34(6):1718–1725, 1995.

[25] V. A. Billock and B. H. Tsou. Hue, saturation and brightness: fundamental properties of color vision derived from dynamic interactions between cortical cell populations. *Journal of Vision*, 2(7):155, November 2002.

[26] D. M. Blei and M. I. Jordan. Modeling annotated data. In *Proceedings of the 26th International ACM Conference on Research and Development in Information Retrieval*, pages 127–134, Toronto, Canada, 2003.

[27] I. Bloch and H. Maitre. Fuzzy mathematical morphologies: a comparative study. *Pattern Recognition*, 28(9):1341–1387, September 1995.

[28] D. Bloomberg and L. Vincent. Pattern matching using the blur hit-miss transform. *Journal of Electronic Imaging*, 9(2):140–150, 2000.

[29] P. Brodatz. *Textures: a photographic album for artists and designers*. Dover, New York, 1966.

[30] H. Burkhardt and S. Siggelkow. Invariant features for discriminating between equivalence classes. In I. Pitas et al., editor, *Nonlinear model-based image video processing and analysis*. John Wiley and Sons, 2000.

[31] C. Busch and M. Eberle. Morphological operations for color-coded images. In *Proceedings of EUROGRAPHICS'95, Computer Graphics Forum*, volume 14, pages 193–204, Maastricht, Netherlands, August 1995.

[32] T. Carron. *Segmentations d'Images Couleur dans la Base Teinte-Luminance-Saturation: Approche Numérique et Symbolique*. PhD thesis, Université de Savoie, France, 1995.

[33] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, August 2002.

[34] E. Y. Chang, K. Goh, G. Sychay, and G. Wu. CBSA: Content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38, 2003.

[35] S. F. Chang, W. Chen, and H. Sundaram. Semantic visual templates: linking visual features to semantics. In *Proceedings of the IEEE Conference on Image Processing*, pages 531–535, Chicago, USA, 1998.

[36] S. K. Chang and A. Hsu. Image information systems: where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431–442, 1992.

[37] J. Chanussot. *Approches Vectorielles ou Marginales pour le Traitement d'Images Multi-composantes*. PhD thesis, Université de Savoie, France, 1998.

[38] J. Chanussot and P. Lambert. Total ordering based on space filling curves for multivalued morphology. In *Proceedings of the International Symposium on Mathematical Morphology (ISMM'98)*, pages 51–58, Amsterdam, Netherlands, June 1998.

[39] J. Chanussot and P. Lambert. Watershed approaches for color image segmentation. In *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, pages 129–133, Antalya, Turkey, June 1999.

[40] J. Chanussot and P. Lambert. Extending mathematical morphology to color image processing. In *International Conference on Color in Graphics and Image Processing*, pages 158–163, Saint-Etienne, France, October 2000.

[41] Y. Chen, J. Li, and J. Z. Wang. *Machine Learning and Statistical Modeling Approaches to Image Retrieval*. Information Retrieval Series. Kluwer Academic Publishers, Dordrecht, 2004.

[42] Y. Chen and J. Z. Wang. Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, 5:913–939, 2004.

[43] Y. Chen, J. Z. Wang, and R. Krovetz. Clue: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing*, 14(8):1187–1201, 2005.

[44] Y. Chen, X. S. Zhou, and T. S. Huang. One-class SVM for learning in image retrieval. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages 34–37, Thessaloniki, Greece, July 2001.

[45] D. Chetverikov. Fundamental structural features in the visual world. In *Proceedings of the International Workshop on Fundamental Structural Properties in Image and Pattern Analysis*, volume 1, pages 47–58, Budapest, Hungary, 1999.

[46] W. W. Chu, C. C. Hsu, A. F. Cardenas, and R. K. Taira. Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering*, 10(6):872–888, November 1998.

[47] CIE. *Proceedings of the Commission International de l'Eclairage*. Cambridge University Press, Cambridge, 1931.

[48] CIE. *International Lighting Vocabulary*. Number 17.4. Cambridge University Press, 4th edition, 1987.

[49] L. Cinque, S. Levialdi, and A. Pellicano. Color-based image retrieval using spatial-chromatic histograms. *IEEE Multimedia Systems*, 2:969–973, 1999.

[50] C. Colombo, A. Rizzi, and I. Genovesi. Histogram families for colour-based retrieval in image databases. In *Proceedings of the International Conference in Image Analysis and Processing*, pages 204 – 211, Florence, Italy, September 1997.

[51] M. Comer and E. Delp. Morphological operations for color image processing. *Journal of Electronic Imaging*, 8(3):279–289, July 1999.

[52] Corel Corporation. Corel gallery images. http://www.corel.com.

[53] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos. The Bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments. *IEEE Transactions on Image Processing*, 9(1):20–37, 2000.

[54] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. Technical Report CSE 06-009, Penn State University, January 2006.

[55] J. G. Daugman. Complete discrete 2d Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(7):1169–1179, July 1998.

[56] A. Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, San Francisco, 1999.

[57] T. Q. Deng and H. J. A. M. Heijmans. Grey-scale morphology based on fuzzy logic. *Journal of Mathematical Imaging and Vision*, 16(2):155–171, March 2002.

[58] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin. An efficient color representation for image retrieval. *IEEE Transactions on Image Processing*, 10(1):140–147, 2001.

[59] Y. Doh, J. Kim, J. Kim, S. Kim, and M. Alam. New morphological detection algorithm based on the hit-miss transform. *Optical Engineering*, 41(1):26–31, 2002.

[60] A. Dong and B. Bhanu. Active concept learning for image retrieval in dynamic databases. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, volume 1, pages 90–95, Nice, France, October 2003.

[61] M. C. d'Ornellas, R. v.d. Boomgaard, and J. Geusebroek. Morphological algorithms for color images based on a generic programming approach. In *Proceedings of the Brazilian Conference on Image Processing and Computer Graphics*, pages 220–228, Rio de Janeiro, Brazil, October 1998.

[62] J. Dowe. Content-based retrieval in multimedia imaging. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, pages 164–167, February 1993.

[63] M. S. Drew, J. Wei, and Z. N. Li. Illumination-invariant image retrieval and video segmentation. *Pattern Recognition*, 32(8):1369–1388, 1999.

[64] G. Durand, C. Thienot, and P. Faudemay. Extraction of composite visual objects from audiovisual materials. In S. Panchanathan, S. F. Chang, and C. C. J. Kuo, editors, *Proceedings of the SPIE Conference on Multimedia Storage and Archiving Systems IV*, volume 3846, pages 194–203, Boston, USA, September 1999.

[65] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the 7th European Conference on Computer Vision*, volume 2353, pages 97–112, Copenhagen, Denmark, May 2002.

[66] J. P. Eakins. Towards intelligent image retrieval. *Pattern Recognition*, 35(1):3–14, 2002.

[67] H. Eidenberger. A new perspective on visual information retrieval. In *Proceedings of the SPIE Electronic Imaging Symposium*, volume 5304, San Jose, USA, January 2004.

[68] A. N. Evans. Vector area morphology for motion field smoothing and interpretation. *IEE Proceedings Vision Image and Signal Processing*, 150(4):219–226, August 2003.

[69] M. Everingham and A.Zisserman et al. The 2005 pascal visual object classes challenge. In *Proceedings of the Machine learning challenges, evaluating predictive uncertainty, visual object classification and recognizing textual entailment (Pascal Workshop 05)*, volume 3944 of *Lecture notes in Artificial Intelligence*, pages 117–176, Southampton, UK, 2005.

[70] M. H. F. Wilkinson F. Tushabe. Content-based image retrieval using shape-size pattern spectra. In A. Nardi and C. Peters, editors, *Proceedings of the CLEF 2007 Workshop*, Budapest, Hungary, 2007.

[71] M. D. Fairchild. *Color Appearance Models*. Addison-Wesley, 1998.

[72] D. Feng, W. C. Siu, and H. J. Zhang, editors. *Multimedia Information Retrieval and Management : Technological Fundamentals and Applications*. Springer, 1st edition, 2003.

[73] G. D. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38(2):179–190, February 2005.

[74] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, September 1995.

[75] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, Massachusetts, 1990.

[76] D. A. Forsyth, J. Malik, M. M. Fleck, H. Greenspan, T. K. Leung, S. Belongie, C. Carson, and C. Bregler. Finding pictures of objects in large collections of images. In *Proceedings of the International Workshop on Object Representation in Computer Vision*, volume 1144 of *Lecture notes in computer science*, pages 335–360, Cambridge, UK, 1996.

[77] J. Fournier and M. Cord. Long-term similarity learning in content-based image retrieval. In *Proceedings of the IEEE International Conference in Image Processing*, volume 1, pages 441–444, New York, USA, September 2002.

[78] G. Fricout and D. Jeulin. Texture classification from their morphological properties. In *Proceedings of the International Conference on Metrology and Properties of Engineering*, pages 42–55, Halmstad University, Sweden, 2003.

[79] B. Gao, T. Y. Liu, T. Qin, X. Zheng, Q. S. Cheng, and W. Y. Ma. Web image clustering by consistent utilization of visual features and surrounding texts. In *Proceedings of the ACM International Conference on Multimedia*, pages 112–121, Singapore, November 2005.

[80] K. Gegenfurtner and D. Kiper. Color vision. *Annual Review of Neuroscience*, 26:181–206, 2003.

[81] S. Gibson, J. A. Bangham, and R. Harvey. Evaluating a colour morphological scale-space. In *Proceedings of the British Machine Vision Conference*, Norwhich, UK, July 2003.

[82] K. S. Goh, E. Y. Chang, and W. C. Lai. Multimodal concept-dependent active learning for image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 564–571, New York, USA, October 2004.

[83] C. Gomila and F. Meyer. Levelings in vector spaces. In *Proceedings of the IEEE Conference on Image Processing*, Kobe, Japan, October 1999.

[84] Google search engine. http://www.google.com.

[85] J. Goutsias and H. J. A. M. Heijmans. Fundamenta morphologicae mathematicae. *Fundamenta Informaticae*, 41(1-2):1–31, January 2000.

[86] J. Goutsias, H. J. A. M. Heijmans, and K. Sivakumar. Morphological operators for image sequences. *Computer Vision and Image Understanding*, 62(3):326–346, November 1995.

[87] K. Grauman and T. Darrell. Benchmarking of image features for content-based image retrieval. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1458–1465, Beijing, China, October 2005.

[88] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.

[89] C. Gu. Multivalued morphology and its application in moving object segmentation and tracking. In P. Maragos, R. W. Schafer, and M. A. Butt, editors, *Mathematical Morphology and its Application to Image and Signal Processing*, pages 345–352. Kluwer Academic Publishers, Boston, USA, 1996.

[90] V. N. Gudivada and V. V. Raghavan. Content-based image retrieval systems. *IEEE Computer*, 28(9):18–22, September 1995.

[91] A. Gupta and R. Jain. Visual information retrieval. *Communications of the ACM*, 40(5):71–79, May 1997.

[92] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar. Multiresolution histograms and their use for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):831–847, 2004.

[93] J. Hafner, H. S. Sawhney, W. Equit, M. Flickner, and W. Niblack. Efficient colour histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, 1995.

[94] J. Han, K. N. Ngan, M. Li, and H. J. Zhang. A memory learning framework for effective image retrieval. *IEEE Transactions on Image Processing*, 14(4):511–524, 2005.

[95] A. Hanbury. *Morphologie Mathématique sur le Cercle Unité: avec Applications aux Teintes et aux Textures Orientées.* PhD thesis, Ecole des Mines, Paris, France, 2002.

[96] A. Hanbury. Circular statistics for colour images. In *8th Computer Vision Winter Workshop*, Valtice, Czech Republic, February 2003.

[97] A. Hanbury. The morphological top-hat operator generalised to multi-channel images. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 672–675, Cambridge, UK, August 2004.

[98] A. Hanbury. Analysis of keywords used in image understanding tasks. In *Proceedings of the OntoImage International Workshop*, Genova, Italy, May 2006.

[99] A. Hanbury. Constructing cylindrical coordinate colour models. *Pattern Recognition Letters*, 4(29):494–500, March 2008.

[100] A. Hanbury. A survey of methods for image annotation. *Journal of Visual Languages & Computing*, January 2008. In press.

[101] A. Hanbury, U. Kandaswamy, and D. A. Adjeroh. Illumination-invariant morphological texture classification. In C. Ronse, L. Najman, and E. Decencière, editors, *Mathematical Morphology: 40 Years On*, volume 30 of *Computational Imaging and Vision*, pages 377–386. Springer-Verlag, Dordrecht, 2005.

[102] A. Hanbury and J. Serra. Mathematical morphology in the HLS colour space. In T. Cootes and C. Taylor, editors, *12th British Machine Vision Conference*, pages 451–460, Manchester, UK, September 2001.

[103] A. Hanbury and J. Serra. Morphological operators on the unit circle. *IEEE Transactions on Image Processing*, 10(12):1842–1850, December 2001.

[104] A. Hanbury and J. Serra. Mathematical morphology in the CIELAB space. *Image Analysis and Stereology*, 21(3):201–206, March 2002.

[105] A. Hanbury and J. Serra. Colour image analysis in 3d-polar coordinates. In *International Conference on Image Processing and its Applications*, Magdeburg, Germany, September 2003.

[106] R. C. Hardie and G. R. Arce. Ranking in R^p and its use in multivariate image estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1(2):197–209, June 1991.

[107] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, Manchester, UK, 1988.

[108] X. He, W. Y. Ma, and H. J. Zhang. Learning an image manifold for retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 17–23, New York, USA, October 2004.

[109] H. Heijmans and J. Serra. Convergence, continuity and iteration in mathematical morphology. *Journal of Visual Communication and Image Representation*, 3(1):84–102, 1992.

[110] H. J. A. M. Heijmans. *Morphological Image Operators*. Advances in Electronics and Electron Physics Series. Academic Press, Boston, 1994.

[111] H. J. A. M. Heijmans and R. Keshet. Inf-semilattice approach to self-dual morphology. *Journal of Mathematical Imaging and Vision*, 17(1):55–80, 2002.

[112] H. J. A. M. Heijmans and C. Ronse. The algebraic basis of mathematical morphology, part I: dilations and erosions. *Computer Vision, Graphics and Image Processing*, 50(3):245–295, June 1990.

[113] C. H. Hoi and M. R. Lyu. Group-based relevance feedback with support vector machine ensembles. In *Proceedings of the International Conference on Pattern Recognition*, pages 874–877, Cambridge, United Kingdom, August 2004.

[114] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, February 1962.

[115] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih. Image indexing using color correlogram. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 762–768, San Juan,Puerto Rico, June 1997.

[116] T. S. Huang, S. Mehrotra, and K. Ramchandran. Multimedia analysis and retrieval system (MARS) project. In *Proceedings of the 33rd Annual Clinic on Library Application of Data Processing*, March 1996.

[117] F. Idris and S. Panchanathan. Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation*, 8(2):146–166, 1997.

[118] Imageclef. http://ir.shef.ac.uk/imageclef/.

[119] E. Jacobs, A. Finkelstein, and D. H. Salesin. Fast multiresolution image querying. In *Proceedings of the 22nd Conference on Computer graphics and interactive techniques*, pages 277 – 286, August 1995.

[120] A. K. Jain and F. Farroknia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[121] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th international ACM conference on Research and development in information retrieval*, pages 119–126, Toronto, Canada, 2003.

[122] S. Jeong, C. S. Won, and R. M. Gray. Image retrieval using color histograms generated by Gauss mixture vector quantization. *Computer Vision and Image Understanding*, 9(1-3):44–66, 2004.

[123] R. Jin, J. Y. Chai, and L. Si. Effective automatic image annotation via a coherent language model and active learning. In *Proceedings of the ACM International Conference on Multimedia*, pages 892–899, New York, USA, October 2004.

[124] Y. Jin, L. Khan, L. Wang, and M. Awad. Image annotations by combining multiple evidence & wordnet. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 706–715, Singapore, November 2005.

[125] F. Jing, M. Li, H. J. Zhang, and B. Zhang. Relevance feedback in region-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):672–681, 2004.

[126] R. Jones and H. Talbot. Morphological filtering for colour images with no new colours. In *Proceedings of the Image and Vision Computing New Zealand (IVCNZ'96)*, pages 149–154, Lower Hutt, New Zealand, August 1996.

[127] D. B. Judd and G. Wyszecki. *Color in business, science and industry.* Wiley series in pure and applied optics. Wiley interscience, New York, 3rd edition, 1975.

[128] K. Karantzalos, D. Argialas, and N. Paragios. Comparing morphological levelings constrained by different markers. In U. M. Braga-Neto G. J. F. Bannon, J. Barrera, editor, *ISMM 2007, 8th International Symposium on Mathematical Morphology*, volume 1, pages 113–124, Rio de Janeiro, Brazil, October 2007. INPE.

[129] T. Kaster, M. Pfeiffer, and C. Bauckhage. Combining speech and haptics for intuitive and efficient navigation through image databases. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 180–187, Vancouver, Canada, 2003.

[130] T. Kato. Database architecture for content-based image retrieval. In *Proceedings of SPIE Image Storage and Retrieval Systems*, pages 112–123, San Jose, USA, February 1992.

[131] R. Keshet. Extensions of morphological operations to complete semilattices and its applications to image and video processing. In H. J. A. M. Heijmans and J. B. T. M. Roerdink, editors, *Mathematical Morphology and its Application to Image and Signal Processing*, pages 35–42, Dordrecht, Germany, 1998. Kluwer Academic Publishers.

[132] R. Keshet. Mathematical morphology on complete semilattices and its applications to image processing. *Fundamenta Informaticae*, 41(1/2):33–56, 2000.

[133] M. L. Kherfi, D. Ziou, and A. Bernardi. Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Computing Surveys*, 36(1):35–67, 2004.

[134] M. Khosravi and R. Schaefer. Template matching based on a greyscale hit-or-miss transform. *IEEE Transactions on Image Processing*, 5(6):1060–1066, 1996.

[135] M. Köppen, Ch. Nowack, and G. Rösel. Pareto-morphology for color image processing. In *Proceedings of the 11th Scandinavian Conference on Image Analysis*, volume 1, pages 195–202, Kangerlussuaq, Greenland, 1999.

[136] M. Koskela. *Interactive Image Retrieval Using Self-Organizing Maps.* PhD thesis, Helsinki University of Technology, Espoo, Finland, November 2003.

[137] L. Kotoulas and I. Andreadis. Colour histogram content-based image retrieval and hardware implementation. *IEE Proceedings - Circuits, Devices and Systems*, 150:387–393, 2003.

[138] J. Kovac, P. Peer, and F. Solina. Human skin color clustering for face detection. In *Proceedings of the IEEE EUROCON*, volume 2, pages 144–148, Ljubljana, Slovenia, September 2003.

[139] J. Laaksonen, M. Koskela, and E. Oja. Picsom-self-organizing image retrieval with MPEG-7 content descriptors. *IEEE Transactions on Neural Networks*, 13(4):841–853, July 2002.

[140] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *Proceedings of the 28th international ACM conference on Research and development in information retrieval*, pages 119–126, Salvador, Brazil, 2005.

[141] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2169–2178, New York, USA, June 2006.

[142] A. Ledda and W. Philips. Majority ordering for colour mathematical morphology. In *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, 2005.

[143] H. Y. Lee, H. K. Lee, and Y. H. Ha. Spatial color descriptor for image retrieval and video segmentation. *IEEE Transactions on Multimedia*, 5(3):358–367, 2003.

[144] S. Lefèvre, J. Holler, and N. Vincent. A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. *Real-Time Imaging*, 9(1):73–98, February 2003.

[145] S. Lefèvre, J. Weber, and D. Sheeren. Automatic building extraction in VHR images using advanced morphological operators. In *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Paris, France, April 2007.

[146] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19, February 2006.

[147] O. Lezoray, C. Meurie, and A. Elmoataz. A graph approach to color mathematical morphology. In *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, Athens, Greece, 2005.

[148] B. Li, K. S. Goh, and E. Y. Chang. Confidence-based dynamic ensemble for image annotation and semantics discovery. In *Proceedings of the ACM International Conference on Multimedia*, pages 195–206, Berkeley, USA, 2003.

[149] F. F. Li, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In *Proceedings of the IEEE CVPR Workshop on Generative Model Based Vision*, volume 12, pages 178–178, Washington, USA, 2004.

[150] J. Li and Y. Li. Multivariate mathematical morphology based on principal component analysis: initial results in building extraction. In *Proceedings of the 20th ISPRS*, volume 35, pages 1168–1173, Istanbul, Turkey, July 2004.

[151] J. Li and J. Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088, 2003.

[152] Y. Y. Lin, T. L. Liu, and H. T. Chen. Semantic manifold learning for image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 249–258, Singapore, November 2005.

[153] F. Liu and R. W. Picard. Periodicity, directionality and randomness: Wold features for image modelling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):722–733, July 1996.

[154] W. Liu and X. Tang. Learning an image-word embedding for image auto-annotation on the nonlinear latent space. In *Proceedings of the ACM International Conference on Multimedia*, pages 127–134, Singapore, November 2005.

[155] G. Louverdis, I. Andreadis, and P. Tsalides. Morphological granulometries for color images. In *2nd Hellenic Conference on Artificial Intelligence*, pages 333–342, Thessaloniki, Greece, April 2002.

[156] G. Louverdis, I. Andreadis, and P. Tsalides. A new fuzzy model for morphological colour image processing. *IEE Proceedings - Vision, Image and Signal Processing*, 149(3):129–139, June 2002.

[157] G. Louverdis, M. Vardavoulia, I. Andreadis, and P. Tsalides. A new approach to morphological color image processing. *Pattern Recognition*, 35(8):1733–1741, September 2002.

[158] R. Lukac, B. Smolka, K. Martin, K. N. Plataniotis, and A. N. Venetsanopoulos. Vector filtering for color imaging. *IEEE Signal Processing Magazine*, 22(1):74–86, January 2005.

[159] Q. T. Luong. Terragalleria. http://www.terragalleria.com.

[160] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. In *Proceedings of the International Conference on Image Processing*, pages 568–571, Santa Barbara, USA, October 1997.

[161] W. Y. Ma and H. J. Zhang. Benchmarking of image features for content-based image retrieval. In *Proceedings of the 32nd Asilomar Conference on Signals, Systems & Computers*, volume 1, pages 253–257, Pacific Grove, USA, November 1998.

[162] T. Maenpaa and M. Pietikainen. Classification with color and texture: jointly or separately? *Pattern Recognition*, 37(8):1629–1640, August 2004.

[163] S. Makrogiannis, G. Economou, S. Fotopoulos, and N. G. Bourbakis. Segmentation of color images using multiscale clustering and graph theoretic region synthesis. *IEEE Transacions on Systems, Man and Cybernetics*, 35(2):224–238, March 2005.

[164] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of large image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[165] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.

[166] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1976.

[167] M. D. Marsicoi, L. Cinque, , and S. Levialdi. Indexing pictorial documents by their content: a survey of current techniques. *Image and Vision Computing*, 15(2):119–141, February 1997.

[168] A. Martinez and J. R. Serra. Semantic access to a database of images: an approach to object-related image retrieval. In *Proceedings of the IEEE Multimedia Systems*, pages 624–629, Florence, Italy, 1999.

[169] Marvel. http://www.research.ibm.com/marvel/.

[170] G. Matheron. *Eléments pour une Théorie des Milieux Poreux*. Masson, Paris, 1967.

[171] G. Matheron. *Random Sets and Integral Geometry*. Wiley, New York, 1975.

[172] E. Mathias and A. Conci. Comparing the influence of color spaces and metrics in content-based image retrieval. In *Proceedings of the International Symposium on Computer Graphics, Image Processing and Vision*, pages 371–378, Rio de Janeiro, Brazil, October 1998.

[173] A. Mauricio and C. Figueirido. Texture analysis of grey-tone images by mathematical morphology: a nondestructive tool for the quantitative assessment of stone decay. *Mathematical Geology*, 32(5):619–642, 2000.

[174] F. Meyer. Levelings, image simplification filters for segmentation. *Journal of Mathematical Imaging and Vision*, 20:59–72, 2004.

[175] F. Meyer and P. Maragos. Nonlinear scale space representation with morphological levelings. *Journal of Visual Communication and Image Representation*, 11(2):245–265, June 2000.

[176] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[177] T. Minka and R. W. Picard. Interactive learning with a society of models. *Pattern Recognition*, 30(4):565–581, April 1997.

[178] A. Mojsilović, J. Kovačević, D. Kall, R. J. Safranek, and S. K. Ganapathy. The vocabulary and grammar of color patterns. *IEEE Transactions on Image Processing*, 9(3):417–431, 2000.

[179] A. Mojsilović and E. Soljanin. Quantization of color spaces and processing of color images by Fibonacci lattices. In *Proceedings of Human Vision and Electronic Imaging V*, volume SPIE 3959, pages 345–355, San Jose, USA, January 2000.

[180] F. Monay and D. Gatica-Perez. On image auto-annotation with latent space models. In *Proceedings of the 11th ACM international conference on Multimedia*, pages 275–278, Berkeley, USA, 2003.

[181] J. Montagnat, H. Duque, J. M. Pierson, V. Breton, L. Brunie, and I. E. Magnin. Medical image content-based queries using the grid. In *Proceedings of the first European HealthGrid Conference*, pages 142–151, Lyon, France, January 2003.

[182] F. C. Montecillo-Puente, V. Ayala-Ramirez, A. Perez-Garcia, and R. E. Sanchez-Yanez. Fuzzy color tracking for robotic tasks. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, volume 3, pages 2769–2773, Washington D.C., USA, October 2003.

[183] H. Muller, S. Marchand-Maillet, and T. Pun. The truth about corel - evaluation in image retrieval. In *Proceedings of the International Conference on Image and Video Retrieval*, volume 2383, pages 36–45, London, United Kingdom, July 2002.

[184] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun. Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters*, 22(5):593–601, 2001.

[185] A. H. Munsell. *A Grammar of Color*. Van Nostrand-Reinhold, New York, 1969.

[186] B. Naegel, N. Passat, and C. Ronse. Grey-level hit-or-miss transforms - part I: Unified theory. *Pattern Recognition*, 40(2):635–647, February 2007.

[187] B. Naegel, N. Passat, and C. Ronse. Grey-level hit-or-miss transforms - part II: Applications to angiographic image processing. *Pattern Recognition*, 40(2):648–658, February 2007.

[188] C. Nastar, M. Mitschke, C. Meilhac, and N. Boujemaa. Surfimage: A flexible content-based image retrieval system. In *Proceedings of the ACM International Conference on Multimedia*, pages 339–344, Bristol, United Kingdom, September 1998.

[189] S. Nene, S. Nayar, and H. Murase. Columbia object image library:coil, 1996.

[190] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: Quering images by content using color, texture, and shape. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, pages 173–187, San Jose, USA, February 1993.

[191] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen. Outex: New framework for empirical evaluation of texture analysis algorithms. In *Proceedings of the 16th ICPR*, volume 1, pages 701–706, Quebec City, Canada, August 2002.

[192] T. Ojala, T. Maenpaa, J. Viertola, J. Kyllonen, and M. Pietikainen. Empirical evaluation of MPEG-7 xml color descriptors in content-based retrieval of semantic image categories. In *Proceedings of the 16th International Conference on Pattern Recognition*, pages 1021–1024, Quebec, Canada, October 2002.

[193] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, January 1996.

[194] T. Ojala, M. Rautiainen, E. Matinmikko, and M. Aittola. Semantic image retrieval with HSV correlograms. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, pages 621–627, Bergen, Norway, 2001.

[195] A. Opelt and A. Pinz. Object localization with boosting and weak supervision for generic object recognition. In *Proceedings of the 14th Scandinavian Conference on Image Analysis*, pages 862–871, Joensuu, Finland, June 2005.

[196] F. Ortiz. Colour mathematical morphology for neural image analysis. *Real-Time Imaging*, 8(5):455–465, December 2002.

[197] F. Ortiz and F. Torres. Vectorial morphological reconstruction for brightness elimination in colour images. *Real-Time Imaging*, 10(6):379–387, December 2004.

[198] F. Ortiz, F. Torres, J. Angulo, and S. Puente. Comparative study of vectorial morphological operations in different color spaces. *Proceedings of Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision*, 4572:259–268, November 2001.

[199] F. Ortiz, F. Torres, and P. Gil. Gaussian noise elimination in colour images by vector-connected filters. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 807–810, Cambridge, UK, August 2004.

[200] F. Ortiz, F. Torres, P. Gil, J. Pomares, S. Puente, and F. Candelas. Ordering by distance for HSI mathematical morphology. In *Proceedings of the 9th Spanish Symposium on Pattern Recognition and Image Analysis*, volume 2, pages 379–384, Castellón, Spain, February 2001.

[201] R. Öten and R. J. P. De Figueiredo. Adaptive alpha-trimmed mean filters under deviations from assumed noise model. *IEEE Transactions on Image Processing*, 13(5):627–639, May 2004.

[202] T. H. Painter, J. Dozier, D. A. Roberts, R. E. Davis, and R. O. Green. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, 85(1):64–77, 2003.

[203] H. Palus. *Colour Spaces*. Chapmann and Hall, 1998.

[204] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. In *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision*, pages 96–102, Sarasota, USA, December 1996.

[205] G. Pass and R. Zabih. Comparing images using joint histograms. *Multimedia Systems*, 7(3):234–240, May 1999.

[206] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, June 1996.

[207] A. Peters. Mathematical morphology for angle-valued images. In *Proceedings of SPIE, Nonlinear Image Processing VIII*, volume 3026, pages 84–94, April 1997.

[208] R. W. Picard. Toward a visual thesaurus. In *Proceedings of the Final Workshop on Multimedia Information Retrieval*, Workshops in Computing, Glasgow, United Kingdom, September 1995.

[209] S. Pinker. *How the Mind Works.* Penguin Press Science, London, 1997.

[210] I. Pitas and P. Tsakalides. Multivariate ordering in color image filtering. *IEEE Transactions on Circuits and Systems for Video Technology*, 1(3):247–259, September 1991.

[211] A. Plaza, P. Martinez, R. Perez, and J. Plaza. A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles. *Pattern Recognition*, 37(6):1097–1116, June 2004.

[212] M. Popovic. Texture analysis using 2d wavelet transform: Theory and applications. In *Proceedings of the 4th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, volume 1, pages 149–158, Nis, Serbia, October 1999.

[213] C. Poynton. Frequently asked questions about color, 1999. http://www.poynton.com.

[214] C. Poynton. *Digital Video and HDTV: Algorithms and Interfaces.* Morgan Kaufmann Publishers, San Francisco, 2003.

[215] I. Pratikakis, I. Vanhamel, H. Sahli, B. Gatos, and S. Perantonis. Watershed-driven region-based image retrieval. In C. Ronse, L. Najman, and E. Decencière, editors, *Mathematical Morphology: 40 years on*, volume 30 of *Computational Imaging and Vision*, pages 207–216. Springer-Verlag, Dordrecht, 2005.

[216] The state hermitage museum. http://www.hermitagemuseum.org/.

[217] T. Quack, U. Monich, L. Thiele, and B. S. Manjunath. Cortina: A system for large scale, content-based web image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 508–511, New York, USA, October 2004.

[218] B. Raducanu and M. Grana. A grayscale hit-or-miss transform based on levet sets. In *Proceedings of the 7th International Conference on Image Processing*, volume 2, pages 931–933, Vancouver, Canada, September 2000.

[219] A. Rao, R. K. Srihari, and Z. Zhang. Spatial color histograms for content-based image retrieval. In *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, pages 183–186, Chicago, USA, November 1999.

[220] A. R. Rao. *A Taxonomy for Texture Description and Identification.* Springer-Verlag, 1990.

[221] A. R. Rao and G. L. Lohse. Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing*, 55(3):218–233, May 1993.

[222] J. Richefeu and A. Manzanera. A morphological dominant points detection and its cellular implementation. In *Proceedings of the IEEE International Symposium on Signal Processing and Its Applications*, volume 2, pages 181–184, Paris, France, July 2003.

[223] J. Rivest. Morphological operators on complex signals. *Signal Processing*, 84(1):133–139, January 2004.

[224] J. Rivest. Granulometries and pattern spectra for radar signals. *Signal Processing*, 86(5):1094–1103, 2006.

[225] C. Ronse. Why mathematical morphology needs complete lattices. *Signal Processing*, 21(2):129–154, October 1990.

[226] C. Ronse. A lattice-theoretical morphological view on template extraction in images. *Journal of Visual Communication and Image Representation*, 7(3):273–295, 1996.

[227] C. Ronse. Reconstructing masks from markers in non-distributive lattices. *Applicable Algebra in Engineering, Communication and Computing*, 19(1):51–85, 2008.

[228] C. Ronse and V. Agnus. Morphology on label images: Flat-type operators and connections. *Journal of Mathematical Imaging and Vision*, 22(2/3):283–307, November 2005.

[229] C. Ronse and H. J. A. M. Heijmans. The algebraic basis of mathematical morphology, part II: openings and closings. *Computer Vision, Graphics and Image Processing*, 54(1):74–97, July 1991.

[230] N. C. Rowe. Marie-4: A high-recall, self-improving web crawler that finds images using captions. *IEEE Intelligent Systems*, 17(4):8–14, 2002.

[231] Y. Rui and T. S. Huang. Optimizing learning in image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 236–245, South Carolina, USA, June 2000.

[232] Y. Rui, T. S. Huang, and S. F. Chang. Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(1):1–23, March 1999.

[233] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in MARS. In *Proceedings of the International Conference on Image Processing*, pages 815–818, Santa Barbara, USA, October 1997.

[234] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Transactions Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[235] L. J. Sartor and A. R. Weeks. Morphological operations on color images. *Journal of Electronic Imaging*, 10(2):548–559, April 2001.

[236] S. Scarlogg, L. Taycher, and M. La Cascia. Image rover: A content-based image browser for the world wide web. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 10–18, San Juan, Puerto Rico, June 1997.

[237] R. Schaefer and D. Casasent. Nonlinear optical hit-miss transform for detection. *Applied Optics*, 34(20):3869–3882, 1995.

[238] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[239] J. Serra. Patent: "Automatic scanning device for analyzing textures" July 2, 1965a, (23 273), France. Patented in : Belgium, Canada, Great-Britain, Japan, Sweden, USA.

[240] J. Serra. *Image Analysis and Mathematical Morphology Vol I.* Academic Press, London, 1982.

[241] J. Serra. *Image Analysis and Mathematical Morphology Vol II: Theoretical Advances.* Academic Press, London, 1988.

[242] J. Serra. Anamorphoses and function lattices. In E. R. Dougherty, editor, *Mathematical Morphology in Image Processing*, chapter 13, pages 483–523. Marcel Dekker, New York, 1993.

[243] J. Serra. Espaces couleur et traitement d'images. Technical Report Internal Note N-34/02/MM, Ecole des Mines de Paris, October 2002.

[244] J. Serra. Morphological segmentations of colour images. In C. Ronse, L. Najman, and E. Decencière, editors, *Mathematical Morphology: 40 Years On*, volume 30 of *Computational Imaging and Vision*, pages 151–176. Springer-Verlag, Dordrecht, 2005.

[245] A. F. Smeaton, P. Over, and W. Kraaij. Trecvid: evaluating the effectiveness of information retrieval tasks on digital video. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 652–655, New York, USA, 2004.

[246] A. W. M. Smeulders, T. Gevers, J. M. Geusebroek, and M. Worring. Invariance in content-based retrieval. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 2, pages 675–678, New York, USA, July 2000.

[247] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[248] J. R. Smith and S. F. Chang. VisualSEEk: A fully automated content-based image query system. In *Proceedings of the 4th ACM international conference on Multimedia*, pages 87–98, Boston, USA, November 1996.

[249] J. R. Smith and S. F. Chang. Visually searching the web for content. *IEEE Multimedia*, 4(3):12–20, July 1997.

[250] P. Soille. Applications of morphological operators. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 3, chapter 12, pages 283–296. Academic Press, San Diego, 1999.

[251] P. Soille. On morphological operators based on rank filters. *Pattern Recognition*, 35(2):527–535, 2002.

[252] P. Soille. *Morphological Image Analysis : Principles and Applications.* Springer-Verlag, Berlin, 2nd edition, 2003.

[253] P. Soille and M. Pesaresi. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9):2042–2055, September 2002.

[254] D. M. Squire, W. Muller, H. Muller, and Thierry Pun. Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters*, 21:1193–1198, 2001.

[255] E. Stringa, A. Teschioni, and C. S. Regazzoni. A classical morphological approach to color image filtering based on space filling surves. In *Proceedings of the IEEE-Eurasip Workshop on Nonlinear Signal and Image Processing*, pages 351–354, Antalya, Turkey, June 1999.

[256] J. Sun, X. Zhang, J. Cui, and L. Zhou. Image retrieval based on colour distribution entropy. *Pattern Recognition Letters*, 27(10):1122–1126, 2006.

[257] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, November 1991.

[258] H. Talbot, C. Evans, and R. Jones. Complete ordering and multivariate mathematical morphology. In H. Heijmans and J. Roerdink, editors, *Mathematical Morphology and Its Applications to Image and Signal Processing*, pages 27–34, Amsterdam, 1998. Kluwer Academic Press.

[259] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6):460–472, June 1978.

[260] K. Tang, J. Astola, and Y. Neuvo. Multivariate ordering and filtering with application to color image processing. In *Proceedings of the 6th European Signal Processing Conference*, pages 1481–1484, Brussels, Belgium, August 1992.

[261] D. M. Titterington. Estimation of correlation coefficients by ellipsoid trimming. *Applied Statistics*, 27(3):227–234, 1978.

[262] S. Tong and E. Tang. Support vector machine active learning for image retrieval. In *Proceedings of the 9th ACM International Conference on Multimedia*, pages 107–118, Ottawa, Canada, 2001.

[263] P. E. Trahanias and A. N. Venetsanopoulos. Color edge detection based on multivariate ordering. In *SPIE Conference, Visual Communications and Image Processing*, pages 1396–1407, Boston, USA, November 1992.

[264] P. E. Trahanias and A. N. Venetsanopoulos. Multichannel image processing using vector-angle ranking. In E. R. Dougherty, J. Astola, and H. G. Longbotham, editors, *SPIE Conference, Nonlinear Image Processing IV*, volume 1902, pages 116–127, San Jose, USA, February 1993.

[265] L. V. Tran and R. Lenz. Compact colour descriptors for colour-based image retrieval. *Signal Processing*, 85(2):233–246, February 2005.

[266] A. Trémeau, C. Fernandez-Maloigne, and P. Bonton. *Image Numérique Couleur*. Dunod, Paris, 2004.

[267] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson. Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):272–285, February 2007.

[268] M. T. Ureikat and H. M. Al-Otum. Angle-valued Mahalanobis-based image morphological operators. In *Proceedings of the 4th WSEAS International Conference on Signal, Speech and Image Processing*, Izmir, Turkey, September 2004.

[269] C. Vachier. Morphological scale-space analysis and feature extraction. In *Proceedings of the International Conference on Image Processing*, volume 3, pages 676–679, Thessaloniki, Greece, October 2001.

[270] A. Vailaya and A. K. Jain. Incremental learning for Bayesian classification of images. In *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, 1999.

[271] A. Vailaya and A. K. Jain. Detecting sky and vegetation in outdoor images. In *Proceedings of the SPIE Conference on Storage and Retrieval for Media Databases VIII*, volume 3972, pages 411–420, San Jose, USA, January 2000.

[272] M. I. Vardavoulia, I. Andreadis, and P. Tsalides. A new vector median filter for colour image processing. *Pattern Recognition Letters*, 22(6-7):675–689, May 2001.

[273] M. I. Vardavoulia, I. Andreadis, and P. Tsalides. Vector ordering and morphological operations for colour image processing: fundamentals and applications. *Pattern Analysis and Applications*, 5(3):271–287, 2002.

[274] Final results of the caltech 256 challenge. http://www.vision.caltech.edu/CaltechChallenge2007/results/.

[275] N. Vasconcelos and M. Kunt. Content-based retrieval from image databases: current solutions and future directions. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 6–9, Thessaloniki, Greece, October 2001.

[276] N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, page 63, South Carolina, USA, June 2000.

[277] R. C. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. In M. Lew, editor, *Principles of Visual Information Retrieval*. Springer, 2001.

[278] R. C. Veltkamp and M. Tanase. Content-based image retrieval systems: A survey. Technical Report UU-CS-2000-34, Utrecht University, 2000.

[279] C. Vertan, M. Malciu, V. Buzuloiu, and V. Popescu. Median filtering techniques for vector valued signals. In *Proceedings of the 3rd IEEE International Conference on Image Processing*, volume 1, pages 977–980, Lausanne, Switzerland, September 1996.

[280] C. Vertan, V. Popescu, and V. Buzuloiu. Morphological like operators for color images. In *Proceedings of the 8th European Signal Processing Conference*, pages 165–168, Trieste, Italy, September 1996.

[281] L. Vincent. Morphological grayscale reconstruction in image analysis:applications and efficient algorithms. *IEEE Transaction on Image Processing*, 2(2):176–201, april 1993.

[282] Viper. http://viper.unige.ch/.

[283] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, Vienna, Austria, 2004.

[284] H. D. Wactlar. Extracting and visualizing knowledge from film and video archives. In *Proceedings of the 2nd International Conference on Knowledge Management I-KNOW*, Graz, Austria, July 2002.

[285] X. J. Wang, W. Y. Ma, Q. C. He, and X. Li. Grouping web image search result. In *Proceedings of the ACM International Conference on Multimedia*, pages 436–439, New York, USA, October 2004.

[286] J. Weber and S. Lefèvre. A multivariate hit-or-miss transform for conjoint spatial and spectral template matching. In *Proceedings of the IEEE International Conference on Image and Signal Processing*, Cherbourg, France, July 2008.

[287] K. R. Weber and S. T. Acton. On connected filters in color image processing. *Journal of Electronic Imaging*, 13(3):619–629, January 2004.

[288] R. Weiss, A. Duda, and D. K. Gifford. Composition and search with a video algebra. *IEEE Multimedia*, 2(1):12–25, 1995.

[289] M. Wheeler and M. A. Zmuda. Processing color and complex data using mathematical morphology. In *IEEE Proceedings of the National Aerospace and Electronics Conference*, volume 1, pages 618–624, Dayton, USA, October 2000.

[290] D. A. White and R. Jain. Similarity indexing: Algorithms and performance. In I. K. Sethi and R. C. Jain, editors, *Proceedings of SPIE Conference On Storage and Retrieval for Still Image and Video Databases IV*, volume 2670, pages 62–73, San Diego, USA, March 1996.

[291] M. H. F. Wilkinson. Generalized pattern spectra sensitive to spatial information. In *Proceedings of the 16th ICPR*, volume 1, pages 21–24, Quebec City, Canada, August 2002.

[292] S. Wilson. Vector morphology and iconic neural networks. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1636–1644, 1989.

[293] S. S. Wilson. Theory of matrix morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):636–652, June 1992.

[294] H. Wu, H. Lu, and S. Ma. Willhunter: Interactive image retrieval with multilevel relevance measurement. In *Proceedings of the International Conference on Pattern Recognition*, volume 2, pages 1009–1012, Cambridge, United Kigdom, August 2004.

[295] X. Xie, H. Liu, S. Goumaz, and W. Y. Ma. Learning user interest for image browsing on small-form-factor devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 671–680, Oregon, USA, 2005.

[296] C. Yang, M. Dong, and F. Fotouhi. Region based image annotation through multiple-instance learning. In *Proceedings of the ACM International Conference on Multimedia*, pages 435–438, Singapore, November 2005.

[297] C. Yang, M. Dong, and F. Fotouhi. Semantic feedback for interactive image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 415–418, Singapore, November 2005.

[298] E. Zaharescu, M. Zamfir, and C. Vertan. Color morphology-like operators based on color geometric shape characteristics. In *Proceedings of the International Symposium on Signals, Circuits and Systems*, volume 1, pages 145–148, Bucarest, Romania, July 2003.

[299] H. J. Zhang and L. Wenyin C. Hu. iFind - a system for semantics and feature based image retrieval over internet. In *Proceedings of the ACM International Conference on Multimedia*, pages 477–478, California, USA, October 2000.

[300] L. Zhang, L. Chen, M. Li, and H. J. Zhang. Automated annotation of human faces in family albums. In *Proceedings of the 11th ACM international conference on Multimedia*, pages 355–358, Berkeley, USA, 2003.

[301] D. Zhao and D. Daut. Shape recognition using morphological transformations. In *Proceedings of the 16th International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 2565–2568, Toronto, Canada, 1991.

[302] X. Zheng, D. Cai, X. He, W. Y. Ma, and X. Lin. Locality preserving clustering for image database. In *Proceedings of the ACM International Conference on Multimedia*, pages 885–891, New York, USA, October 2004.

[303] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, 2003.

Résumé

Cette thèse porte principalement sur l'extension de la morphologie mathématique aux images couleur. Les opérateurs morphologiques qui en dérivent sont ensuite appliqués au problème de la recherche d'images par le contenu.

L'étude de ce problème débute avec le choix d'un espace de couleurs "pertinent", et il a été décidé d'utiliser les espaces de couleurs s'appuyant sur un triplet luminance-saturation-teinte, plébiscités pour leur intuitivité et leur bonne représentation du système de la vision humaine. Suite au choix d'espace de couleurs, nos efforts se sont portés sur le seul prérequis pour permettre l'extension de la morphologie mathématique aux données multivariées : un ordre et des opérateurs d'extremums pour des données vectorielles. Plus précisément, nous nous sommes intéressés à l'approche lexicographique, du fait de ses propriétés théoriques intéressantes et de sa capacité d'adaptation à différents contextes applicatifs. Plusieurs solutions ont été proposées pour s'affranchir de son principal inconvénient qui consiste en la prioritisation extrême de la première bande spectrale ou couleur. En outre, une attention particulière a été portée au traitement de la teinte, susceptible de poser des problèmes si elle est ordonnée comme une valeur scalaire alors que sa nature est périodique. Par conséquent nous présentons une approche utilisant plusieurs teintes de référence. Cette partie se termine par l'extension aux images multivariées de la transformée en tout-ou-rien, découlant des résultats obtenus précédemment.

La deuxième partie de ce travail concerne l'application de la morphologie couleur au problème de la recherche d'images par le contenu. Dans ce but, nous avons introduit plusieurs descripteurs globaux de couleur et de texture. Plus précisément, nous nous sommes intéressés au concept d'histogramme multi-échelles et avons élaboré deux méthodes permettant leur intégration avec les opérateurs morphologiques. En outre, des granulométries spécifiques aux images couleur ont aussi été étudiées. Pour la description de texture, nous avons combiné les deux principaux outils morphologiques de caractérisation de texture, la granulométrie et la covariance. Finalement, nous avons proposé l'architecture d'un système pour la description, l'indexation et la recherche par le contenu à base de mots-clés, qui fournit une solution flexible en termes d'adaptabilité aux différents niveaux d'hétérogénéité de collections d'images.

Mot-clés: analyse d'images couleur, morphologie mathématique multivariée, ordres vectoriels, description d'images, extraction de caractéristiques, recherche d'images par le contenu.