
Image Search Engines

An Overview

by

Th. Gevers and A.W.M. Smeulders
(`{gevers, smeulders}@science.uva.nl`)

Faculty of Science
University of Amsterdam
1098 SJ Amsterdam, The Netherlands

CONTENTS

PREFACE	vii
1 IMAGE SEARCH ENGINES: AN OVERVIEW	1
1.1 Overview of the chapter	1
1.2 Image Domains	6
1.2.1 Search modes	6
1.2.2 The sensory gap	8
1.2.3 The semantic gap	9
1.2.4 Discussion	10
1.3 Image Features	11
1.3.1 Color	11
1.3.2 Shape	15
1.3.3 Texture	15
1.3.4 Discussion	18
1.4 Representation and Indexing	18
1.4.1 Grouping data	18
1.4.2 Features accumulation	20
1.4.3 Feature accumulation and image partitioning	22
1.4.4 Salient features	23
1.4.5 Shape and object features	24
1.4.6 Structure and lay-out	25
1.4.7 Discussion	26
1.5 Similarity and Search	27
1.5.1 Semantic interpretation	27
1.5.2 Similarity between features	27
1.5.3 Similarity of object outlines	30
	v

1.5.4	Similarity of object arrangements	31
1.5.5	Similarity of salient features	31
1.5.6	Discussion	32
1.6	Interaction and Learning	33
1.6.1	Interaction on a semantic level	33
1.6.2	Classification on a semantic level	33
1.6.3	Learning	34
1.6.4	Discussion	34
1.7	Conclusion	34
	BIBLIOGRAPHY	36

PREFACE

TG/AS
University of Amsterdam
June, 2003

IMAGE SEARCH ENGINES: AN OVERVIEW

In this chapter, we present an overview on the theory, techniques and applications of content-based image retrieval. We choose patterns of use, image domains and computation as the pivotal building blocks of our survey. A graphical overview of the content-based image retrieval scheme is given in Fig. 1.1. Derived from this scheme, we follow the data as they flow through the computational process, see Fig. 1.3, with the conventions indicated in Fig. 1.2. In all of this chapter, we follow the review in [155] closely.

We focus on still images and leave video retrieval as a separate topic. Video retrieval could be considered as a broader topic than image retrieval as video is more than a set of isolated images. However, video retrieval could also be considered to be simpler than image retrieval since, in addition to pictorial information, video contains supplementary information such as motion, and spatial and time constraints e.g. video disclose its objects more easily as many points corresponding to one object move together and are spatially coherent in time. In still pictures the user's narrative expression of intention is in image selection, object description and composition. Video, in addition, has the linear time line as an important information cue to assist the narrative structure.

1.1 Overview of the chapter

The overview of the basic components, to be discussed in this chapter, is given in Fig. 1.1 and the corresponding dataflow process is shown in Fig. 1.3. The sections in this chapter harmonize with the data as they flow from one computational component to another as follows:

- *Interactive query formulation*: Interactive query formulation is offered either by query (sub)image(s) or by offering a pattern of feature values and weights. To achieve interactive query formulation, an image is sketched, recorded or selected from an image repository. With the query formulation, the aim to search for particular images in the database. The mode of search might be one of the following three

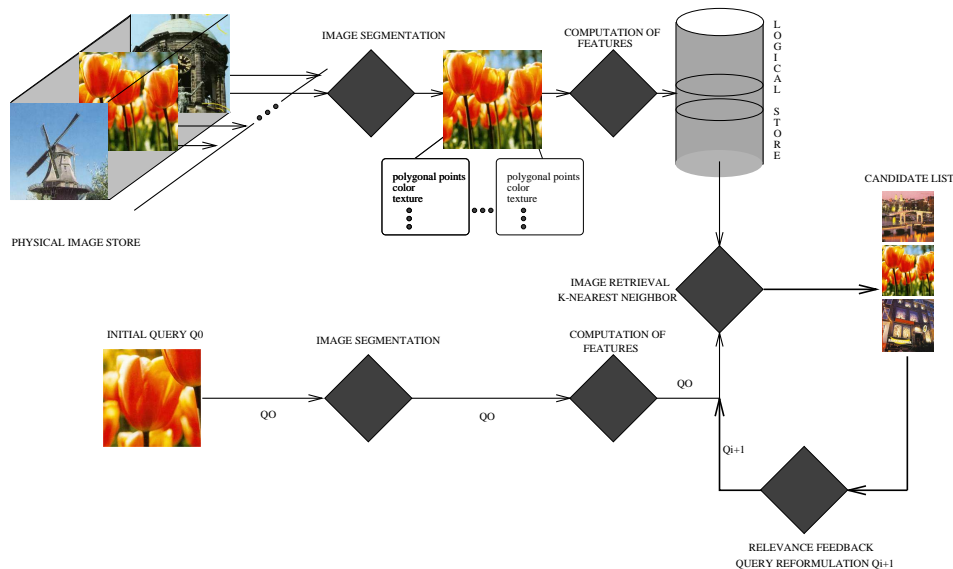


Figure 1.1. Overview of the basic concepts of the content-based image retrieval scheme as considered in this chapter. First, features are extracted from the images in the database which are stored and indexed. This is done off-line. The on-line image retrieval process consists of a query example image from which image features are extracted. These image features are used to find the images in the database which are most similar. Then, a candidate list of most similar images is shown to the user. From the user feedback the query is optimized and used as a new query in an iterative manner.

categories: *search by association*, *target search*, and *category search*. For search by association, the intention of the user is to browse through a large collection of images without a specific aim. Search by association tries to find interesting images and is often applied in an iterative way by means of relevance feedback. Target search is to find similar (target) images in the image database. Note that "similar image" may imply a (partially) identical image, or a (partially) identical object in the image. The third class is category search, where the aim is to retrieve an arbitrary image which is typical for a specific class or genre (e.g. indoor images, portraits, city views). As many image retrieval systems are assembled around one of these three search modes, it is important to get more insight in these categories and their structure. Search modes will be discussed in Section 1.2.1.

- *Image domains*: The definition of image features depends on the repertoire of images under consideration. This repertoire can be ordered along the complexity of variations imposed by the imaging conditions such as illumination and viewing geometry going from *narrow domains* to *broad domains*. For images from a narrow









Arrow and symbol conventions in this paper			
Arrow	Domain	Element	Description
	I	$i(\mathbf{x})$	Image field
	T	$t(\mathbf{x})$	Segmented image field
	F	\mathbf{f}	Feature vector
	F	\mathbf{f}	Saliency feature
	H	\mathbf{h}	Hierarchically ordered feature set
	Z	z	Interpretation
	S	s	Similarity
			Control

Figure 1.2. Data flow and symbol conventions as used in this chapter. Different styles of arrows indicate different data structures.

domain there will be a restricted variability of their pictorial content. Examples of narrow domains are stamp collections and face databases. For *broad domains*, images may be taken from objects from unknown viewpoints and illumination. For example, two recordings taken from the same object from different viewpoints will yield different shadowing, shading and highlighting cues changing the intensity data fields considerably. Moreover, large differences in the illumination color will drastically change the photometric content of images even when they are taken from the same scene. Hence, images from *broad domains* have a large pictorial variety which is called the sensory gap to be discussed in Section 1.2.2. Furthermore, low-level image features are often too restricted to describe images on a conceptual or semantic level. This semantic gap is a well-known problem in content-based image retrieval and will be discussed in Section 1.2.3.

•*Image features*: Image feature extraction is an important step for image indexing and search. Image feature extraction modules should take into account whether the image domain is narrow or broad. In fact, they should consider to which of the imaging conditions they should be invariant to such a change in viewpoint, object pose, and illumination. Further, image features should be concise and complete and

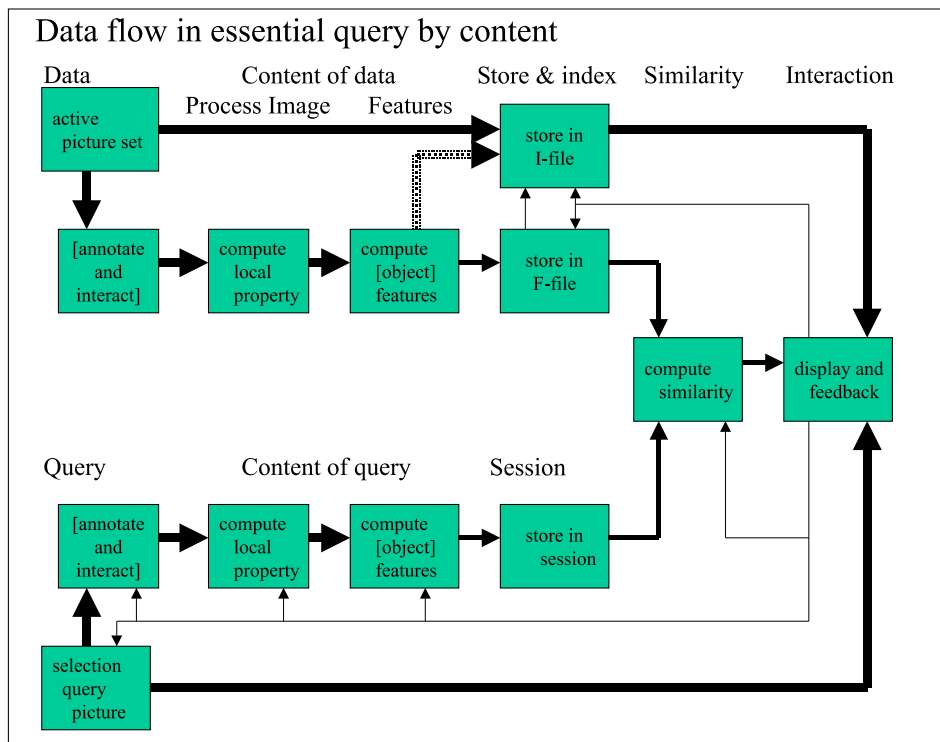


Figure 1.3. Basic algorithmic components of query by pictorial example captured in a data-flow scheme while using the conventions of Fig. 1.2.

at the same having high discriminative power. In general, a tradeoff exists between the amount of invariance and selectivity. In Section 1.3, a taxonomy on feature extraction modules is given from an image processing perspective. The taxonomy can be used to select the proper feature extraction method for a specific application based on whether images come from broad domains and which search goals are at hand (target/category/associate search). In Section 1.3.1, we first focus on color content descriptors derived from image processing technology. Various color based image search methods will be discussed based on different representation schemes such as color histograms, color moments, color edge orientation, and color correlograms. These image representation schemes are created on the basis of RGB , and other color systems such as HSI and $CIE L^*a^*b^*$. For example, the $L^*a^*b^*$ space has been designed to conform to the human perception of color similarity. If the appreciation of a human observer of an object is based on the perception of certain conspicuous items in the image, it is natural to direct the computation of broad domain features to these points and regions. Similarly, a biologically plausible ar-

chitecture [84] of center-surround processing units is likely to select regions which humans would also focus on first. Further, color models are discussed which are robust to a change in viewing direction, object geometry and illumination. Image processing for shape is outlined in Section 1.3.2. We focus on *local shape* which are image descriptors capturing salient details in images. Finally, in Section 1.3.3, our attention is directed towards texture and a review is given on texture features describing local color characteristics and their spatial layout.

- *Representation and indexing*

Representation and indexing will be discussed in Section 1.4. In general, the image feature set is represented by vector space, probabilistic or logical models. For example, for the vector space model, weights can be assigned corresponding to the feature frequency giving the well-known histogram form. Further, for accurate image search, it is often desirable to assign weights in accordance to the importance of the image features. The image feature weights used for both images and queries can be computed as the product of the features frequency multiplied by the inverse collection frequency factor. In this way, features are emphasized having high feature frequencies but low overall collection frequencies. More on feature accumulation and representation is discussed in Section 1.4.2. In addition to feature representation, indexing is required to speed up the search process. Indexing techniques include adaptive histogram binning, signature files, and hashing. Further, tree-based indexing schemes have been developed for indexing the stored images so that similar images can be identified efficiently at some additional costs in memory, such as a k-d tree, R*-tree or a SS-tree, [69] for example.

Throughout the chapter, a distinction is made between *weak* and *strong* segmentation. Weak segmentation is a local grouping approach usually focusing on conspicuous regions such as edges, corners and higher-order junctions. In Section 1.4.4, various methods are discussed to achieve weak segmentation. Strong segmentation is the extraction of the complete contour of an object in an image. Obviously, strong segmentation is far more difficult than weak segmentation and is hard to achieve if not impossible for broad domains.

- *Similarity and search*

The actual matching process can be seen as a search for images in the stored image set closest to the query specification. As both the query and the image data set is captured in feature form, the similarity function operates between the weighted feature sets. To make the query effective, close attention has to be paid to the selection of the similarity function. A proper similarity function should be robust to object fragmentation, occlusion and clutter by the presence of other objects in the view. For example, it is known that the mean square and the Euclidean similarity measure provides accurate retrieval without any object clutter [59] [162]. A detailed overview on similarity and search is given in Section 1.5.

- *Interaction and Learning*

Visualization of the feature matching results gives the user insight in the importance of the different features. Windowing and information display techniques can be used to establish communications between system and user. In particular,

new visualization techniques such as 3D virtual image clouds can be used to designate certain images as relevant to the user's requirements. These relevant images are then further used by the system to construct subsequent (improved) queries. Relevance feedback is an automatic process designed to produce improved query formulations following an initial retrieval operation. Relevance feedback is needed for image retrieval where users find it difficult to formulate pictorial queries. For example, without any specific query image example, the user might find it difficult to formulate a query (e.g. to retrieve an image of a car) by image sketch or by offering a pattern of feature values and weights. This suggests that the first search is performed by an initial query formulation and a (new) improved query formulation is constructed based on the search results with the goal to retrieve more relevant images in the next search operations. Hence, from the user feedback giving negative/positive answers, the method can automatically learn which image features are more important. The system uses the feature weighting given by the user to find the images in the image database which are optimal with respect to the feature weighting. For example, the *search by association* allows users to refine iteratively the query definition, the similarity or the examples with which the search was started. Therefore, systems in this category are highly interactive. Interaction, relevance feedback and learning are discussed in Section 1.6.

- *Testing*

In general, image search systems are assessed in terms of precision, recall, query-processing time as well as reliability of a negative answer. Further, the relevance feedback method is assessed in terms of the number of iterations to approach to the ground-truth. Today, more and more images are archived yielding a very large range of complex pictorial information. In fact, the average number of images, used for experimentation as reported in the literature, augmented from a few in 1995 to over a hundred thousand by now. It is important that the dataset should have ground-truths i.e. images which are (non) relevant to a given query. In general, it is hard to get these ground-truths. Especially for very large datasets. A discussion on system performance is given in Section 1.6.

1.2 Image Domains

In this section, we discuss patterns in image search applications, the repertoire of images, the influence of the image formation process, and the semantic gap between image descriptors and the user.

1.2.1 Search modes

We distinguish three broad categories of search modes when using a content-based image retrieval system, see Fig. 1.4.

- There is a broad variety of methods and systems designed to browse through a large set of images from unspecified sources, which is called *search by association*. At the start, users of search by association have no specific aims other than to

Target-, category- and association-search in image retrieval			
	<i>Target</i>	<i>Category</i>	<i>Association</i>
<i>Object goal</i>	1 specific object	an arbitrary object from 1 specific class	not defined at start
<i>Query by example</i>	1 ... N objects	1 ... N objects with class labels	N objects plus association
<i>Similarity</i>	feature-based	class driven	session-specific
<i>Events in F-space</i>	proximity to query	class membership	clusters
<i>Feedback</i>	rank ordered on proximity	likelihood on class membership	relevance feedback on association value
<i>Interactive update:</i>			
<i>of images of query</i>	-	expand query	refine on the way
<i>of features of query</i>	refine on the way	refine on the way	alter on the way
<i>of similarity measure</i>	-	adapt to group	reshape to goal

Figure 1.4. Three patterns in the purpose of content-based retrieval systems.

find interesting images. Search by association often implies iterative refinement of the search, the similarity or the examples with which the search was initiated. Systems in this category are highly interactive, where the query specification may be defined by sketch [28] or by example images. The oldest realistic example of such a system is probably [91]. The result of the search can be manipulated interactively by relevance feedback [76]. To support the quest for relevant results, also other sources than images are employed, for example [163].

- Another class of search mode is *target search* with the purpose to find a specific image. The search may be for a precise copy of the image in mind, as in searching art catalogues, e.g. [47]. Target search may also be for another image of the same object the user has an image of. This is target search by example. Target search may also be applied when the user has a specific image in mind and the target is interactively specified as similar to a group of given examples, for instance [29]. These systems are suited to search for stamps, paintings, industrial components, textile patterns, and catalogues in general.

- The third class of search modes is *category search*, aiming at retrieving an arbitrary image representative for a specific class. This is the case when the user has an example and the search is for other elements of the same class or genre. Categories may be derived from labels or may emerge from the database [164], [105]. In category search, the user may have available a group of images and the search is for additional images of the same class [25]. A typical application of category search is catalogues of varieties. In [82], [88], systems are designed for classifying

trademarks. Systems in this category are usually interactive with a domain specific definition of similarity.

1.2.2 The sensory gap

In the repertoire of images under consideration (the image domain) there is a gradual distinction between narrow and broad domains [154]. At one end of the spectrum, we have the narrow domain:

A narrow domain has a limited and predictable variability in all relevant aspects of its appearance.

Hence, in a narrow domain one finds images with a reduced diversity in their pictorial content. Usually, the image formation process is similar for all recordings. When the object's appearance has limited variability, the semantic description of the image is generally well-defined and largely unique. An example of a narrow domain is a set of frontal views of faces, recorded against a clear background. Although each face is unique and has large variability in the visual details, there are obvious geometrical, physical and illumination constraints governing the pictorial domain. The domain would be wider in case the faces had been photographed from a crowd or from an outdoor scene. In that case, variations in illumination, clutter in the scene, occlusion and viewpoint will have a major impact on the analysis.

On the other end of the spectrum, we have the broad domain:

A broad domain has an unlimited and unpredictable variability in its appearance even for the same semantic meaning.

In broad domains images are polysemic, and their semantics are described only partially. It might be the case that there are conspicuous objects in the scene for which the object class is unknown, or even that the interpretation of the scene is not unique. The broadest class available today is the set of images available on the Internet.

Many problems of practical interest have an image domain in between these extreme ends of the spectrum. The notions of broad and narrow are helpful in characterizing patterns of use, in selecting features, and in designing systems. In a broad image domain, the gap between the feature description and the semantic interpretation is generally wide. For narrow, specialized image domains, the gap between features and their semantic interpretation is usually smaller, so domain-specific models may be of help.

For broad image domains in particular, one has to resort to generally valid principles. Is the illumination of the domain white or colored? Does it assume fully visible objects, or may the scene contain clutter and occluded objects as well? Is it a 2D-recording of a 2D-scene or a 2D-recording of a 3D-scene? The given characteristics of illumination, presence or absence of occlusion, clutter, and differences in camera viewpoint, determine the demands on the methods of retrieval.

The sensory gap is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.

The sensory gap makes the description of objects an ill-posed problem: it yields uncertainty in what is known about the state of the object. The sensory gap is particularly poignant when a precise knowledge of the recording conditions is missing. The 2D-records of different 3D-objects can be identical. Without further knowledge, one has to decide that they *might* represent the same object. Also, a 2D-recording of a 3D- scene contains information accidental for that scene and that sensing but one does not know what part of the information is scene related. The uncertainty due to the sensory gap does not only hold for the viewpoint, but also for occlusion (where essential parts telling two objects apart may be out of sight), clutter, and illumination.

1.2.3 The semantic gap

As stated in the previous sections, content-based image retrieval relies on multiple low-level features (e.g. color, shape and texture) describing the image content. To cope with the sensory gap, these low-level features should be consistent and invariant to remain representative for the repertoire of images in the database. For image retrieval by query by example, the on-line image retrieval process consists of a query example image, given by the user on input, from which low-level image features are extracted. These image features are used to find images in the database which are most similar to the query image. A drawback, however, is that these low-level image features are often too restricted to describe images on a conceptual or semantic level. It is our opinion that ignoring the existence of the semantic gap is the cause of many disappointments on the performance of early image retrieval systems.

The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

A user wants to search for images on a conceptual level e.g. images containing particular objects (target search) or conveying a certain message or genre (category search). Image descriptions, on the other hand, are derived by low-level data-driven methods. The *semantic* search by the user and the low-level *syntactic* image descriptors may be disconnected. Association of a complete semantic system to image data would entail, at least, solving the general object recognition problem. Since this problem is yet unsolved and will likely to stay unsolved in its entirety, research is focused on different methods to associate higher level semantics to data-driven observables.

Indeed, the most reasonable tool for semantic image characterization entails annotation by keywords or captions. This converts content-based image access to

(textual) information retrieval [134]. Common objections to the practice of labeling are cost and coverage. On the cost side, labeling thousands of images is a cumbersome and expensive job to the degree that the deployment of the economic balance behind the database is likely to decrease. To solve the problem, systems presented in [140], [139] use a program that explores the Internet collecting images and inserting them in a predefined taxonomy on the basis of the text surrounding them. A similar approach for digital libraries is taken by [19]. On the coverage side, labeling is seldom complete, context sensitive and, in any case, there is a significant fraction of requests whose semantics can't be captured by labeling alone [7], [72]. Both methods will cover the semantic gap only in isolated cases.

1.2.4 Discussion

We have discussed three broad types of search categories: target search, category search and search by association. Target search is related to the classical methods in the field of pattern matching and computer vision such as object recognition and image matching. However, image retrieval differs from traditional pattern matching by considering more and more images in the database. Therefore, new challenges in content-based retrieval are in the huge amount of images to search among, the query specification by multiple images, and in the variability of imaging conditions and object states. Category search connects to statistical pattern recognition methods. However, compared to traditional pattern recognition, new challenges are in the interactive manipulation of results, the usually very large number of object classes, and the absence of an explicit training phase for feature and classifier tuning (active learning). Search by association is the most distant from the classical field of computer vision. It is severely hampered by the semantic gap. As long as the gap is there, use of content-based retrieval for browsing will not be within the grasp of the general public as humans are accustomed to rely on the immediate semantic imprint the moment they see an image.

An important distinction we have discussed is that between broad and narrow domains. The broader the domain, the more browsing or search by association should be considered during system set-up. The narrower the domain, the more target search should be taken as search mode.

The major discrepancy in content-based retrieval is that the user wants to retrieve images on a semantic level, but the image characterizations can only provide similarity on a low-level syntactic level. This is called the semantic gap. Furthermore, another discrepancy is that between the properties in an image and the properties of the object. This is called the sensory gap. Both the semantic and sensory gap play a serious limiting role in the retrieval of images based on their content.

1.3 Image Features

Before starting the discussion on image features, it is important to keep in mind that content-based retrieval does not depend on a complete description of the pictorial content of the image. It is sufficient that a retrieval system presents similar images, i.e. similar in some user defined sense. The description of the content by image features should serve that goal primarily.

One such goal can be met by using invariance as a tool to deal with the accidental distortions in the image content introduced by the sensory gap. From Section 1.2.2, it is clear that invariant features may carry more object-specific information than other features as they are insensitive to the accidental imaging conditions such as illumination, object pose and camera viewpoint. The aim of invariant image features is to identify objects no matter from how and where they are observed at the loss of some of the information content.

Therefore, the degree of invariance, should be tailored to the recording circumstances. In general, a feature with a very wide class of invariance loses the power to discriminate among object differences. The aim is to select the tightest set of invariants suited for the expected set of non-constant conditions. What is needed in image search is a specification of the minimal invariant conditions in the specification of the query. The minimal set of invariant conditions can only be specified by the user as it is part of his or hers intention. For each image retrieval query a proper definition of the desired invariance is in order. Does the applicant wish search for the object in rotation and scale invariance? illumination invariance? viewpoint invariance? occlusion invariance? The oldest work on invariance in computer vision has been done in object recognition as reported among others in [119] for shape and [181] for color. Invariant description in image retrieval is relatively new, but quickly gaining ground, for a good introduction see [15], [30], [57].

1.3.1 Color

Color has been an active area of research in image retrieval, more than in any other branch of computer vision. Color makes the image take values in a color vector space. The choice of a color system is of great importance for the purpose of proper image retrieval. It induces the equivalent classes to the actual retrieval algorithm. However, no color system can be considered as universal, because color can be interpreted and modeled in different ways. Each color system has its own set of color models, which are the parameters of the color system. Color systems have been developed for different purposes: 1. display and printing processes: *RGB*, *CMY*; 2. television and video transmission efficiency: *YIQ*, *YUV*; 3. color standardization: *XYZ*; 4. color uncorrelation: $I_1 I_2 I_3$; 5. color normalization and representation: *rgb*, *xyz*; 6. perceptual uniformity: $U^*V^*W^*$, $L^*a^*b^*$, $L^*u^*v^*$; 7. and intuitive description: *HSI*, *HSV*. With this large variety of color systems, the inevitable question arises which color system to use for which kind of image retrieval application. To this end, criteria are required to classify the various color systems for

the purpose of content-based image retrieval. Firstly, an important criterion is that the color system is independent of the underlying imaging device. This is required when images in the image database are recorded by different imaging devices such as scanners, camera's and camrecorder (e.g. images on Internet). Another prerequisite might be that the color system should exhibit perceptual uniformity meaning that numerical distances within the color space can be related to human perceptual differences. This is important when images are to be retrieved which should be visually similar (e.g. stamps, trademarks and paintings databases). Also, the transformation needed to compute the color system should be linear. A non-linear transformation may introduce instabilities with respect to noise causing poor retrieval accuracy. Further, the color system should be composed of color models which are understandable and intuitive to the user. Moreover, to achieve robust image retrieval, color invariance is an important criterion. In general, images and videos are taken from objects from different viewpoints. Two recordings made of the same object from different viewpoints will yield different shadowing, shading and highlighting cues.

Only when there is no variation in the recording or in the perception than the *RGB* color representation is a good choice. *RGB*-representations are widely in use today. They describe the image in its literal color properties. An image expressed by *RGB* makes most sense when recordings are made in the absence of variance, as is the case, e.g., for art paintings [72], the color composition of photographs [47] and trademarks [88], [39], where two dimensional images are recorded in frontal view under standard illumination conditions.

A significant improvement over the *RGB*-color space (at least for retrieval applications) comes from the use of normalized color representations [162]. This representation has the advantage of suppressing the intensity information and hence is invariant to changes in illumination intensity and object geometry.

Others approaches use the Munsell or the $L^*a^*b^*$ -spaces because of their relative perceptual uniformity. The $L^*a^*b^*$ color system has the property that the closer a point (representing a color) is to another point, the more visual similar the colors are. In other words, the magnitude of the perceived color difference of two colors corresponds to the Euclidean distance between the two colors in the color system. The $L^*a^*b^*$ system is based on the three dimensional coordinate system based on the opponent theory using black-white L^* , red-green a^* , and yellow-blue b^* components. The L^* axis corresponds to the lightness where $L^* = 100$ is white and $L^* = 0$ is black. Further, a^* ranges from red $+a^*$ to green $-a^*$ while b^* ranges from yellow $+b^*$ to blue $-b^*$. The chromaticity coordinates a^* and b^* are insensitive to intensity and has the same invariant properties as normalized color. Care should be taken when digitizing the non-linear conversion to $L^*a^*b^*$ -space [117].

The *HSV*-representation is often selected for its invariant properties. Further, the human color perception is conveniently represented by these color models where I is an attribute in terms of which a light or surface color may be ordered on a scale from dim to bright. S denotes the relative white content of a color and H is the color aspect of a visual impression. The problem of H is that it becomes unstable when S

is near zero due to the non-removable singularities in the nonlinear transformation, which a small perturbation of the input can cause a large jump in the transformed values [62]. H is invariant under the orientation of the object with respect to the illumination intensity and camera direction and hence more suited for object retrieval. However, H is still dependent on the color of the illumination [57].

A wide variety of tight photometric color invariants for object retrieval were derived in [59] from the analysis of the dichromatic reflection model. They derive for matte patches under white light the invariant color space $(\frac{R-G}{R+G}, -\frac{B-R}{B+R}, \frac{G-B}{G+B})$, only dependent on sensor and surface albedo. For a shiny surface and white illumination, they derive the invariant representation as $\frac{|R-G|}{|R-G|+|B-R|+|G-B|}$ and two more permutations. The color models are robust against major viewpoint distortions.

Color constancy is the capability of humans to perceive the same color in the presence of variations in illumination which change the physical spectrum of the perceived light. The problem of color constancy has been the topic of much research in psychology and computer vision. Existing color constancy methods require specific a priori information about the observed scene (e.g. the placement of calibration patches of known spectral reflectance in the scene) which will not be feasible in practical situations, [48], [52], [97] for example. In contrast, without any a priori information, [73], [45] use illumination-invariant moments of color distributions for object recognition. However, these methods are sensitive to object occlusion and cluttering as the moments are defined as an integral property on the object as one. In global methods in general, occluded parts will disturb recognition. [153] circumvents this problem by computing the color features from small object regions instead of the entire object. Further, to avoid sensitivity on object occlusion and cluttering, simple and effective illumination-independent color ratio's have been proposed by [53], [121], [60]. These color constant models are based on the ratio of surface albedos rather than the recovering of the actual surface albedo itself. However, these color models assume that the variation in spectral power distribution of the illumination can be modeled by the coefficient rule or von Kries model, where the change in the illumination color is approximated by a 3x3 diagonal matrix among the sensor bands and is equal to the multiplication of each RGB -color band by an independent scalar factor. The diagonal model of illumination change holds exactly in the case of narrow-band sensors. Although standard video camera's are not equipped with narrow-band filters, spectral sharpening could be applied [46] to achieve this to a large extent.

The color ratio's proposed by [121] are given by: $N(C^{\vec{x}_1}, C^{\vec{x}_2}) = \frac{C^{\vec{x}_1} - C^{\vec{x}_2}}{C^{\vec{x}_2} + C^{\vec{x}_1}}$ and those proposed by [53] are defined by: $F(C^{\vec{x}_1}, C^{\vec{x}_2}) = \frac{C^{\vec{x}_1}}{C^{\vec{x}_2}}$ expressing color ratio's between two neighboring image locations, for $C \in \{R, G, B\}$, where \vec{x}_1 and \vec{x}_2 denote the image locations of the two neighboring pixels.

The color ratio's of [60] are given by: $M(C_1^{\vec{x}_1}, C_1^{\vec{x}_2}, C_2^{\vec{x}_1}, C_2^{\vec{x}_2}) = \frac{C_1^{\vec{x}_1} C_2^{\vec{x}_2}}{C_1^{\vec{x}_2} C_2^{\vec{x}_1}}$ expressing the color ratio between two neighboring image locations, for $C_1, C_2 \in \{R, G, B\}$ where \vec{x}_1 and \vec{x}_2 denote the image locations of the two neighboring pixels. All

these color ratios are device dependent, not perceptual uniform and they become unstable when intensity is near zero. Further, N and F are dependent on the object geometry. M has no viewing and lighting dependencies. In [55] a thorough overview is given on color models for the purpose of image retrieval. Figure 1.5 shows the taxonomy of color models with respect to their characteristics. For more information we refer to [55].

Color system	Device indep.	Perc. Uniform	Linear	Intuitive	View point	Object shape	Highlights	Illum. Intensity	Illum. SPD
RGB	-	-	+	-	-	-	-	-	-
XYZ	+	-	+	-	-	-	-	-	-
Norm. rgb	-	-	-	-	+	+	-	+	-
Norm. xyz	+	-	-	-	+	+	-	+	-
L*a*b*	+	+	-	-	-	-	-	-	-
U*V*W*	+	+	-	-	-	-	-	-	-
111213	-	-	+	-	-	-	-	-	-
YIQ	-	-	+	-	-	-	-	-	-
YUV	-	-	+	-	-	-	-	-	-
Intensity	-	-	+	+	-	-	-	-	-
Hue	-	-	-	+	+	+	+	+	-
Saturation	-	-	-	+	+	+	-	+	-
F, N	-	-	-	-	+	-	-	+	+
M	-	-	-	-	+	+	-	+	+

Figure 1.5. *a. Overview of the dependencies differentiated for the various color systems. + denotes that the condition is satisfied - denotes that the condition is not satisfied.*

Rather than invariant descriptions, another approach to cope with the inequalities in observation due to surface reflection is to search for clusters in a color histogram of the image. In the RGB -histogram, clusters of pixels reflected off an object form elongated streaks. Hence, in [126], a non-parametric cluster algorithm in RGB -space is used to identify which pixels in the image originate from one uniformly colored object.

1.3.2 Shape

Under the name 'local shape' we collect all properties that capture conspicuous geometric details in the image. We prefer the name local shape over other characterization such as differential geometrical properties to denote the result rather than the method.

Local shape characteristics derived from directional color derivatives have been used in [117] to derive perceptually conspicuous details in highly textured patches of diverse materials. A wide, rather unstructured variety of image detectors can be found in [159].

In [61], a scheme is proposed to automatic detect and classify the physical nature of edges in images using reflectance information. To achieve this, a framework is given to compute edges by automatic gradient thresholding. Then, a taxonomy is given on edge types based upon the sensitivity of edges with respect to different imaging variables. A parameter-free edge classifier is provided labeling color transitions into one of the following types: (1) shadow-geometry edges, (2) highlight edges, (3) material edges. In figure 1.6.a, six frames are shown from a standard video often used as a test sequence in the literature. It shows a person against a textured background playing ping-pong. The size of the image is 260x135. The images are of low quality. The frames are clearly contaminated by shadows, shading and inter-reflections. Note that each individual object-parts (i.e. T-shirt, wall and table) is painted homogeneously with a distinct color. Further, that the wall is highly textured. The results of the proposed reflectance based edge classifier are shown in figure 1.6.b-d. For more details see [61].

Combining shape and color both in invariant fashion is a powerful combination as described by [58] where the colors inside and outside affine curvature maximums in color edges are stored to identify objects.

Scale space theory was devised as the complete and unique primary step in pre-attentive vision, capturing all conspicuous information [178]. It provides the theoretical basis for the detection of conspicuous details on any scale. In [109] a series of Gabor filters of different directions and scale have been used to enhance image properties [136]. Conspicuous shape geometric invariants are presented in [135]. A method employing local shape and intensity information for viewpoint and occlusion invariant object retrieval is given in [143]. The method relies on voting among a complete family of differential geometric invariants. Also, [170] searches for differential affine-invariant descriptors. From surface reflection, in [5] the local sign of the Gaussian curvature is computed, while making no assumptions on the albedo or the model of diffuse reflectance.

1.3.3 Texture

In computer vision, texture is considered as all what is left after color and local shape have been considered or it is given in terms of structure and randomness. Many common textures are composed of small textons usually too large in number to be perceived as isolated objects. The elements can be placed more or less regularly

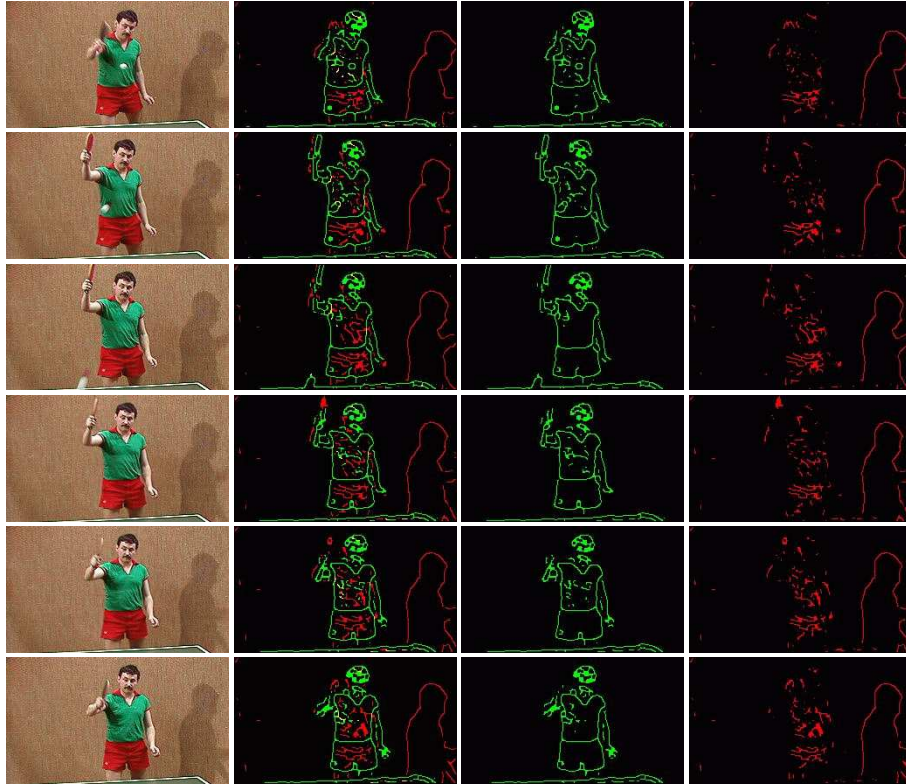


Figure 1.6. *Frames from a video showing a person against a textured background playing ping-pong. From left to right column. a. Original color frame. b. Classified edges. c. Material edges. d. Shadow and geometry edges.*

or randomly. They can be almost identical or subject to large variations in their appearance and pose. In the context of image retrieval, research is mostly directed towards statistical or generative methods for the characterization of patches.

Basic texture properties include the Markovian analysis dating back to Haralick in 1973 and generalized versions thereof [95], [64]. In retrieval, the property is computed in a sliding mask for localization [102], [66].

Another important texture analysis technique uses multi-scale auto-regressive MRSAR-models, which consider texture as the outcome of a deterministic dynamic system subject to state and observation noise [168], [110]. Other models exploit statistical regularities in the texture field [9].

Wavelets [33] have received wide attention. They have often been considered for their locality and their compression efficiency. Many wavelet transforms are generated by groups of dilations or dilations and rotations that have been said to have

some semantic correspondent. The lowest levels of the wavelet transforms [33], [22] have been applied to texture representation [96], [156], sometimes in conjunction with Markovian analysis [21]. Other transforms have also been explored, most notably fractals [41]. A solid comparative study on texture classification from mostly transform-based properties can be found in [133].

When the goal is to retrieve images containing objects having irregular texture organization, the spatial organization of these texture primitives is, in worst case, random. It has been demonstrated that for irregular texture, the comparison of gradient distributions achieves satisfactory accuracy [122], [130] as opposed to fractal or wavelet features. Therefore, most of the work on texture image retrieval is stochastic from nature [12], [124], [190]. However, these methods rely on grey-value information which is very sensitive to the imaging conditions. In [56] the aim is to achieve content-based image retrieval of textured objects in natural scenes under varying illumination and viewing conditions. To achieve this, image retrieval is based on matching feature distributions derived from color invariant gradients. To cope with object cluttering, region-based texture segmentation is applied on the target images prior to the actual image retrieval process. In Figure 1.7 results are shown of color invariant texture segmentation for image retrieval. From the results, we can observe that RGB and normalized color $\theta_1\theta_2$, is highly sensitive to a change in illumination color. Only M is insensitive to a change in illumination color. For more information we refer to [56].

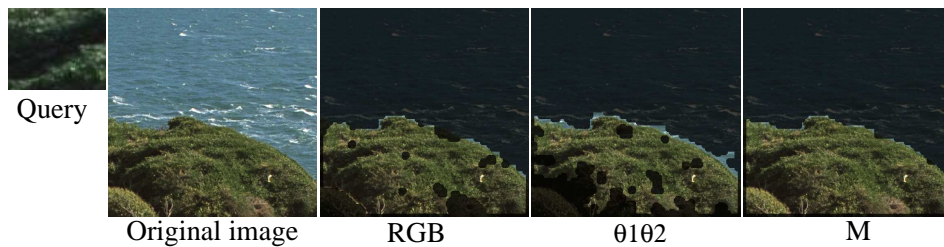


Figure 1.7. *a. Query texture under different illumination b. Target image c. Segmentation result based on RGB. d. Segmentation result based on variant of rgb. e. Segmentation result based on color ratio gradient M.*

Texture search proved also to be useful in satellite images [100] and images of documents [31]. Textures also served as a support feature for segmentation-based recognition [106], but the texture properties discussed so far offer little semantic referent. They are therefore ill-suited for retrieval applications in which the user wants to use verbal descriptions of the image. Therefore, in retrieval research, in [104] the Wold features of periodicity, directionality, and randomness are used, which agree reasonably well with linguistic descriptions of textures as implemented in [127].

1.3.4 Discussion

First of all, image processing in content-based retrieval should primarily be engaged in enhancing the image information of the query, not on describing the content of the image in its entirety.

To enhance the image information, retrieval has set the spotlights on color, as color has a high discriminatory power among objects in a scene, much higher than gray levels. The purpose of most image color processing is to reduce the influence of the accidental conditions of the scene and sensing (i.e. the sensory gap). Progress has been made in tailored color space representation for well-described classes of variant conditions. Also, the application of geometrical description derived from scale space theory will reveal viewpoint and scene independent salient point sets thus opening the way to similarity of images on a few most informative regions or points.

In this chapter, we have made a separation between color, local geometry and texture. At this point it is safe to conclude that the division is an artificial labeling. For example, wavelets say something about the local shape as well as the texture, and so may scale space and local filter strategies do. For the purposes of content-based retrieval an integrated view on color, texture and local geometry is urgently needed as only an integrated view on local properties can provide the means to distinguish among hundreds of thousands different images. A recent advancement in that direction is the fusion of illumination and scale invariant color and texture information into a consistent set of localized properties [74]. Also in [16], homogeneous regions are represented as collections of ellipsoids of uniform color or texture, but invariant texture properties deserve more attention [167] and [177]. Further research is needed in the design of complete sets of image properties with well-described variant conditions which they are capable of handling.

1.4 Representation and Indexing

In the first subsection, we discuss the ultimate form of spatial data by grouping the data into object silhouettes, clusters of points or point-sets. In the next subsection, we leave the spatial domain, to condense the pictorial information into feature values.

1.4.1 Grouping data

In content-based image retrieval, the image is often divided in parts before features are computed from each part. Partitionings of the image aim at obtaining more selective features by selecting pixels in a trade of against having more information in features when no subdivision of the image is used at all. We distinguish the following partitionings:

- When searching for an object, it would be most advantageous to do a complete object segmentation first:

Strong segmentation is a division of the image data into regions in such a way that region T contains the pixels of the silhouette of object O in the real world and nothing else, specified by: $T = O$.

It should be noted immediately that object segmentation for broad domains of general images is not likely to succeed, with a possible exception for sophisticated techniques in very narrow domains.

- The difficulty of achieving strong segmentation may be circumvented by weak segmentation where grouping is based on data- driven properties:

Weak segmentation is a grouping of the image data in conspicuous regions T internally homogeneous according to some criterion, hopefully with $T \subset O$.

The criterion is satisfied if region T is within the bounds of object O , but there is no guarantee that the region covers all of the object's area. When the image contains two nearly identical objects close to each other, the weak segmentation algorithm may falsely observe just one patch. Fortunately, in content-based retrieval, this type of error is rarely obstructive for the goal. In [125], the homogeneity criterion is implemented by requesting that colors be spatially coherent vectors in a region. Color is the criterion in [49], [126]. In [16], [114], the homogeneity criterion is based on color and texture. The limit case of weak segmentation is a set of isolated points [143], [59]. No homogeneity criterion is needed then, but the effectiveness of the isolated points rest on the quality of their selection. When occlusion is present in the image, weak segmentation is the best one can hope for. Weak segmentation is used in many retrieval systems either as a purpose of its own or as a pre-processing stage for data-driven model- based object segmentation.

- When the object has a (nearly) fixed shape, like a traffic light or an eye, we call it a sign:

Localizing signs is finding an object with a fixed shape and semantic meaning, with $T = \mathbf{x}_{\text{center}}$.

Signs are helpful in content based retrieval as they deliver an immediate and unique semantic interpretation.

- The weakest form of grouping is partitioning:

A partitioning is a division of the data array regardless of the data, symbolized by: $T \neq O$.

The area T may be the entire image, or a conventional partitioning as the central part of the image against the upper, right, left and lower parts [75]. The feasibility of fixed partitioning comes from the fact that image are created in accordance with certain canons or normative rules, such as placing the horizon about 2/3 up in the picture, or keeping the main subject in the central area. This rule is often violated, but this violation, in itself, has semantic significance. Another possibility of partitioning is to divide the image in tiles of equal size and summarize the dominant feature values in each tile [129].

1.4.2 Features accumulation

In the computational process given in Fig. 1.3, features are calculated next. The general class of accumulating features aggregate the spatial information of a partitioning irrespective of the image data. A special type of accumulative features are the global features which are calculated from the entire image. F_j (see Fig. 1.2) is the set of accumulative features or a set of accumulative features ranked in a histogram. F_j is part of feature space \mathcal{F} . T_j is the partitioning over which the value of F_j is computed. In the case of global features $T_{j=void}$ represents the image, otherwise T_j represents a fixed tiling of the image. The operator h may hold relative weights, for example to compute transform coefficients.

A simple but very effective approach to accumulating features is to use the histogram, that is the set of features $\mathbf{F}(m)$ ordered by histogram index m .

One of the earlier approaches to color-based image matching, using the color at pixels directly as indices, has been proposed by Swain and Ballard [162]. If the *RGB* or normalized color distributions of two images are globally similar, the matching rate is high. The work by Swain and Ballard has had an enormous impact on color-based histogram matching resulting in many histogram variations.

For example, the QBIC system [42] allows for a user-defined computation of the histogram by the introduction of variable k denoting the number of bins of the histogram. Then, for each $3 \times k$ cells, the average modified Munsell color is computed together with the five most frequently occurring colors. Using a standard clustering algorithm they obtain k super cells resulting in the partitioning of the color system.

In [58] various color invariant features are selected to construct color pattern-cards. First, histograms are created in a standard way. Because the color distributions of histograms depend on the scale of the recorded object (e.g. distance object-camera), they define color pattern-cards as thresholded histograms. In this way, color pattern-cards are scale-independent by indicating whether a particular color model value is substantially present in an image or not. Matching measures are defined, expressing similarity between color pattern-cards, robust to a substantial amount of object occlusion and cluttering. Based on the color pattern-cards and matching functions, a hashing scheme is presented offering run-time image retrieval independent of the number of images in the image database.

In the ImageRover system, proposed by [147], the $L^*u^*v^*$ color space is used where each color axis has been split into 4 equally sized bins resulting in a total of 64 bins. Further, [37] uses the $L^*a^*b^*$ system to compute the average color and covariance matrix of each of the color channels. [158] uses the *HSV* color space to obtain a partition into 144 bins giving more emphasis on hue H than value V and saturation I . Further, [4] also focuses on the *HSV* color space to extract regions of dominant colors. To obtain colors which are perceptually the same but still being distinctive, [165] proposes to partition the *RGB* color space into 220 subspaces. [36] computes the average color describing a cell of a 4×4 grid which is superimposed on the image. [149] uses the $L^*a^*b^*$ color space because the color space consists of perceptually uniform colors, which better matches the human perception of color.

[65] roughly partitions the Munsell color space into eleven color zones. Similar partitioning have been proposed by [29] and [24].

Another approach, proposed by [161], is the introduction of the cumulative color histogram which generate more dense vectors. This enables to cope with coarsely quantized color spaces. [186] proposes a variation of the cumulative histograms by applying cumulative histograms to each sub-space.

Other approaches are based on the computation of moments of each color channel. For example, [6] represents color regions by the first three moments of the color models in the *HSV*-space. Instead of constructing histograms from color invariants, [73], [45] propose the computation of illumination-invariant moments from color histograms. In a similar way, [153] computes the color features from small object regions instead of the entire object.

[85] proposes to use integrated wavelet decomposition. In fact, the color features generate wavelet coefficients together with their energy distribution among channels and quantization layers. Similar approaches based on wavelets have been proposed by [175], [101].

All of this is in favor of the use of histograms. When very large data sets are at stake, plain histogram comparison will saturate the discrimination. For a 64-bin histogram, experiments show that for reasonable conditions, the discriminatory power among images is limited to 25,000 images [160]. To keep up performance, in [125], a joint histogram is used providing discrimination among 250,000 images in their database rendering 80% recall among the best 10 for two shots from the same scene using simple features. Other joint histograms add local texture or local shape [68], directed edges [87], and local higher order structures [47].

Another alternative is to add a dimension representing the local distance. This is the correlogram [80], defined as a 3- dimensional histogram where the colors of any pair are along the first and second dimension and the spatial distance between them along the third. The autocorrelogram defining the distances between pixels of identical colors is found on the diagonal of the correlogram. A more general version is the geometric histogram [1] with the normal histogram, the correlogram and several alternatives as special cases. This also includes the histogram of the triangular pixel values reported to outperform all of the above as it contains more information.

A different view on accumulative features is to demand that all information (or all relevant information) in the image is preserved in the feature values. When the bit-content of the features is less than the original image, this boils down to compression transforms. Many compression transforms are known, but the quest is for transforms simultaneously suited as retrieval features. As proper querying for similarity is based on a suitable distance function between images, the transform has to be applied on a metric space. And, the components of the transform have to correspond to semantically meaningful characteristics of the image. And, finally, the transform should admit indexing in compressed form yielding a big computational advantage over having the image be untransformed first. [144] is just one of many where the cosine-based JPEG-coding scheme is used for image retrieval.

The JPEG-transform fulfills the first and third requirement but fails on a lack of semantics. In the MPEG-standard the possibility to include semantic descriptors in the compression transform is introduced [27]. For an overview of feature indexes in the compressed domain, see [108]. In [96], a wavelet packet was applied to texture images and, for each packet, entropy and energy measures were determined and collected in a feature vector. In [83], vector quantization was applied in the space of coefficients to reduce its dimensionality. This approach was extended to incorporate the metric of the color space in [141]. In [86] a wavelet transform was applied independently to the three channels of a color image, and only the sign of the most significant coefficients is retained. In [3], a scheme is offered for a broad spectrum of invariant descriptors suitable for application on Fourier, wavelets and splines and for geometry and color alike.

1.4.3 Feature accumulation and image partitioning

The lack of spatial information for methods based on feature accumulation might yield lower retrieval accuracy. As for general image databases, a manual segmentation is not feasible due to the sensory gap, a simple approach is to divide images into smaller sub-images and then index them. This is known as fixed partitioning. Other systems use a segmentation scheme, prior to the actual image search, to partition each image into regions. Nearly all region-based partitioning schemes use some kind of weak segmentation decomposing the image into coherent regions rather than complete objects (strong segmentation).

Fixed Partitioning

The simplest way is to use a fixed image decomposition in which an image is partitioned into equally sized segments. The disadvantage of a fixed partitioning is that blocks usually do not correspond with the visual content of an image. For example, [65] splits an image into nine equally sized sub-images, where each sub-region is represented by a color histogram. [67] segments the image by a quadtree, and [99] uses a multi-resolution representation of each image. [36] also uses a 4x4 grid to segment the image. [148] partitions images into three layers, where the first layer is the whole image, the second layer is a 3x3 grid and the third layer a 5x5 grid. A similar approach is proposed by [107] where three levels of a quadtree is used to segment the images. [37] proposes the use of inter-hierarchical distances measuring the difference between color vectors of a region and its sub-segments. [20] uses an augmented color histogram capturing the spatial information between pixels together to the color distribution. In [59] the aim is to combine color and shape invariants for indexing and retrieving images. Color invariant edges are derived from which shape invariant features are computed. Then computational methods are described to combine the color and shape invariants into a unified high-dimensional histogram for discriminatory object retrieval. [81] proposes the use of color correlograms for image retrieval. Color correlograms integrate the spatial information of colors by expressing the probability that a pixel of color c_i lies at a certain distance from a

pixel of color c_j . It is shown that color correlograms are robust to a change in background, occlusion, and scale (camera zoom). [23] introduces the spatial chromatic histograms, where for every pixel the percentage of pixels having the same color is computed. Further, the spatial information is encoded by baricenter of the spatial distribution and the corresponding deviation.

Region-based Partitioning

Segmentation is a computational method to assess the set of points in an image which represent one object in the scene. As discussed before, many different computational techniques exist, none of which is capable of handling any reasonable set of real world images. However, in this case, *weak segmentation* may be sufficient to recognize an object in a scene. Therefore, in [12] an image representation is proposed providing a transformation from the raw pixel data to a small set of image regions which are coherent in color and texture space. This so-called Blob-world representation is based on segmentation using the Expectation-Maximization algorithm on combined color and texture features. In the Picasso system [13], a competitive learning clustering algorithm is used to obtain a multiresolution representation of color regions. In this way, colors are represented in the $l^*u^*v^*$ space through a set of 128 reference colors as obtained by the clustering algorithm. [63] proposes a method based on matching feature distributions derived from color ratio gradients. To cope with object cluttering, region-based texture segmentation is applied on the target images prior to the actual image retrieval process. [26] segments the image first into homogeneous regions by split and merge using a color distribution homogeneity condition. Then, histogram intersection is used to express the degree of similarity between pairs of image regions.

1.4.4 Salient features

As the information of the image is condensed into just a limited number of feature values, the information should be selected with precision for greatest saliency and proven robustness. That is why saliency in [103] is defined as the special points, which survive longest when gradually blurring the image in scale space. Also in [137] lifetime is an important selection criterion for salient points in addition to wiggleness, spatial width, and phase congruency. To enhance the quality of salient descriptions, in [170] invariant and salient features of local patches have been considered. In each case, the image is summarized in a list of conspicuous points. In [143] salient and invariant transitions in gray value images are recorded. Similarly, in [59], [54], photometric invariance is the leading principle in summarizing the image in salient transitions in the image. Salient feature calculations lead to sets of regions or points with known location and feature values capturing their salience.

In [16], first the most conspicuous homogeneous regions in the image are derived and mapped into feature space. Then, expectation-maximization [35] is used to determine the parameters of a mixture of Gaussians to model the distribution of points into the feature space. The means and covariance matrices of these Gaus-

sians, projected on the image plane, are represented as ellipsoids characterized by their center \mathbf{x} , their area, eccentricity, and direction. The average values of the color and texture descriptions inside the ellipse are also stored.

Various color image segmentation methods have been proposed which account for the image formation process, see for instance the work collected by Wolff, Shafer and Healey [181]. [150] presented the dichromatic reflection model, a physical model of reflection which states that two distinct types of reflection - surface and body reflection - occur, and that each type can be decomposed into a relative spectral distribution and a geometric scale factor. [93] developed a color segmentation algorithm based on the dichromatic reflection model. The method is based on evaluating characteristic shapes of clusters in red-green-blue (*RGB*) space followed by segmentation independent of the object's geometry, illumination and highlights. To achieve robust image segmentation, however, surface patches of objects in view must have a rather broad distribution of surface normals which may not hold for objects in general. [10] developed a similar image segmentation method using the *H-S* color space instead of the *RGB*-color space. [73] proposed a method to segment images on the basis of normalized color. However, [92] showed that normalized color and hue are singular at some *RGB* values and unstable at many others.

1.4.5 Shape and object features

The theoretically best way to enhance object-specific information contained in images is by segmenting the object in the image. But, as discussed above, the brittleness of segmentation algorithms prevents the use of automatic segmentation in broad domains. And, in fact, in many cases it is not necessary to know exactly where an object is in the image as long as one can identify the presence of the object by its unique characteristics. When the domain is narrow a tailored segmentation algorithm may be needed more, and fortunately also be better feasible.

The object internal features are largely identical to the accumulative features, now computed over the object area. They need no further discussion here.

An abundant comparison of shape for retrieval can be found in [113], evaluating many features on a 500-element trademark data set. Straightforward features of general applicability include Fourier features and moment invariants of the object this time, sets of consecutive boundary segments, or encoding of contour shapes [40].

For retrieval, we need a shape representation that allows a robust measurement of distances in the presence of considerable deformations. Many sophisticated models widely used in computer vision often prove too brittle for image retrieval. On the other hand, the (interactive) use of retrieval makes some mismatch acceptable and, therefore precision can be traded for robustness and computational efficiency.

More sophisticated methods include elastic matching and multi-resolution representation of shapes. In elastic deformation of image portions [34], [123] or modal matching techniques [145] image patches are deformed to minimize a cost functional that depends on a weighed sum of the mismatch of the two patches and on

the deformation energy. The complexity of the optimization problem depends on the number of points on the contour. Hence, the optimization is computationally expensive and this, in spite of the greater precision of these methods, has limited their diffusion in image databases.

Multi-scale models of contours have been studied as a representation for image databases in [118]. Contours were extracted from images and progressively smoothed by dividing them into regions of constant sign of the second derivative and progressively reducing the number of such regions. At the final step, every contour is reduced to an ellipsoid which could be characterized by some of the features in [47]. A different view on multi-resolution shape is offered in [98], where the contour is sampled by a polygon, and then simplified by removing points from the contour until a polygon survives selecting them on perceptual grounds. When computational efficiency is at stake an approach for the description of the object boundaries is found in [189] where an ordered set of critical points on the boundary are found from curvature extremes. Such sets of selected and ordered contour points are stored in [112] relative to the basis spanned by an arbitrary pair of the points. All point pairs are used as a basis to make the redundant representation geometrically invariant, a technique similar to [182] for unordered point sets.

For retrieval of objects in 2D-images of the 3D-worlds, a viewpoint invariant description of the contour is important. A good review of global shape invariants is given in [138].

1.4.6 Structure and lay-out

When feature calculations are available for different entities in the image, they may be stored with a relationship between them. Such a structural feature set may contain feature values plus spatial relationships, a hierarchically ordered set of feature values, or relationships between point sets or object sets. Structural and layout feature descriptions are captured in a graph, hierarchy or any other ordered set of feature values and their relationships.

To that end, in [111], [49] lay-out descriptions of an object are discussed in the form of a graph of relations between blobs. A similar lay-out description of an image in terms of a graph representing the spatial relations between the objects of interest was used in [128] for the description of medical images. In [51], a graph is formed of topological relationships of homogeneous *RGB*-regions. When selected features and the topological relationships are viewpoint invariant, the description is viewpoint invariant, but the selection of the *RGB*-representation as used in the paper will only suit that purpose to a limited degree. The systems in [78], [157] studies spatial relationships between regions each characterized by locations, size and features. In the later system, matching is based on the 2D-string representation founded by Chang [17]. For a narrow domain, in [128], [132] the relevant element of a medical X-ray image are characterized separately and joined together in a graph that encodes their spatial relations.

Starting from a shape description, the authors in [98] decompose an object into

its main components making the matching between images of the same object easier. Automatic identification of salient regions in the image based on non-parametric clustering followed by decomposition of the shapes found into limbs is explored in [50].

1.4.7 Discussion

General content-based retrieval systems have dealt with segmentation brittleness in a few ways. First, a weaker version of segmentation has been introduced in content-based retrieval. In weak segmentation the result is a homogeneous region by some criterion, but not necessarily covering the complete object silhouette. It results in a fuzzy, blobby description of objects rather than a precise segmentation. Salient features of the weak segments capture the essential information of the object in a nutshell. The extreme form of the weak segmentation is the selection of a salient point set as the ultimately efficient data reduction in the representation of an object, very much like the focus-of-attention algorithms for an earlier age. Only points on the interior of the object can be used for identifying the object, and conspicuous points at the borders of objects have to be ignored. Little work has been done how to make the selection. Weak segmentation and salient features are a typical innovation of content-based retrieval. It is expected that salience will receive much attention in the further expansion of the field especially when computational considerations will gain in importance.

The alternative is to do no segmentation at all. Content-based retrieval has gained from the use of accumulative features, computed on the global image or partitionings thereof disregarding the content, the most notable being the histogram. Where most attention has gone to color histograms, histograms of local geometric properties and texture are following. To compensate for the complete loss of spatial information, recently the geometric histogram was defined with an additional dimension for the spatial layout of pixel properties. As it is a superset of the histogram an improved discriminability for large data sets is anticipated. When accumulative features they are calculated from the central part of a photograph may be very effective in telling them apart by topic but the center does not always reveals the purpose. Likewise, features calculated from the top part of a picture may be effective in telling indoor scenes from outdoor scenes, but again this holds to a limited degree. A danger of accumulative features is their inability to discriminate among different entities and semantic meanings in the image. More work on semantic-driven groupings will increase the power of accumulative descriptors to capture the content of the image.

Structural descriptions match well with weak segmentation, salient regions and weak semantics. One has to be certain that the structure is within one object and not an accidental combination of patches which have no meaning in the object world. The same brittleness of strong segmentation lurks here. We expect a sharp increase in the research of local, partial or fuzzy structural descriptors for the purpose of content-based retrieval especially of broad domains.

1.5 Similarity and Search

When the information from images is captured in a feature set, there are two possibilities for endowing them with meaning: one derives an unilateral interpretation from the feature set or one compares the feature set with the elements in a given data set on the basis of a similarity function.

1.5.1 Semantic interpretation

In content-based retrieval it is useful to push the semantic interpretation of features derived from the image as far as one can.

Semantic features aim at encoding interpretations of the image which may be relevant to the application.

Of course, such interpretations are a subset of the possible interpretations of an image. To that end, consider a feature vector \mathbf{F} derived from an image i . For given semantic interpretations z from the set of all interpretations \mathcal{Z} , a strong semantic feature with interpretation z_j would generate a $P(z|\mathbf{F}) = \delta(z - z_j)$. If the feature carries no semantics, it would generate a distribution $P(z|\mathbf{F}) = P(z)$ independent of the value of the feature. In practice, many feature types will generate a probability distribution that is neither a pulse nor independent of the feature value. This means that the feature value skews the interpretation of the image, but does not determine it completely.

Under the umbrella *weak semantics* we collect the approaches that try to combine features in some semantically meaningful interpretation. Weak semantics aims at encoding in a simple and approximate way a subset of the possible interpretations of an image that are of interest in a given application. As an example, the system in [28] uses color features derived from Itten's color theory to encode the semantics associated to color contrast and harmony in art application.

In the MAVIS2-system [90] data are considered at four semantic levels, embodied in four layers called the raw media, the selection, the selection expression and conceptual layers. Each layer encodes information at an increasingly symbolic level. Agents are trained to create links between features, feature signatures at the selection layer, inter-related signatures at the selection expression layer, and concept (expressed as textual labels) at the conceptual layer. In addition to the vertical connections, the two top layers have intra-layer connections that measure the similarity between concepts at that semantic level and contribute to the determination of the similarity between elements at the lower semantic level.

1.5.2 Similarity between features

A different road to assign a meaning to an observed feature set, is to compare a pair of observations by a similarity function. While searching for a query image $i_q(\mathbf{x})$ among the elements of the data set of images, $i_d(\mathbf{x})$, knowledge of the domain will be expressed by formulating a similarity measure $S_{q,d}$ between the images q and d

on the basis of some feature set. The similarity measure depends on the type of features.

At its best use, the similarity measure can be manipulated to represent different semantic contents; images are then grouped by similarity in such a way that close images are similar with respect to use and purpose. A common assumption is that the similarity between two feature vectors \mathbf{F} can be expressed by a positive, monotonically non increasing function. This assumption is consistent with a class of psychological models of human similarity perception [152], [142], and requires that the feature space be metric. If the feature space is a vector space, d often is a simple Euclidean distance, although there is indication that more complex distance measures might be necessary [142]. This similarity model was well suited for early query by example systems, in which images were ordered by similarity with one example.

A different view sees similarity as an essentially probabilistic concept. This view is rooted in the psychological literature [8], and in the context of content-based retrieval it has been proposed, for example, in [116].

Measuring the distance between histograms has been an active line of research since the early years of content-based retrieval, where histograms can be seen as a set of ordered features. In content-based retrieval, histograms have mostly been used in conjunction with color features, but there is nothing against being used in texture or local geometric properties.

Various distance functions have been proposed. Some of these are general functions such as Euclidean distance and cosine distance. Others are specially designed for image retrieval such as histogram intersection [162]. The Minkowski-form distance for two vectors or histograms \vec{k} and \vec{l} with dimension n is given by:

$$\mathcal{D}_M^k(\vec{k}, \vec{l}) = \left(\sum_{i=1}^n |k_i - l_i|^\rho \right)^{1/\rho} \quad (1.5.1)$$

The Euclidean distance between two vectors \vec{k} and \vec{l} is defined as follows:

$$\mathcal{D}_E(\vec{k}, \vec{l}) = \sqrt{\sum_{i=1}^n (k_i - l_i)^2} \quad (1.5.2)$$

The Euclidean distance is an instance of the Minkowski distance with $k = 2$.

The cosine distance compares the feature vectors of two images and returns the cosine of the angle between the two vectors:

$$\mathcal{D}_C(\vec{k}, \vec{l}) = 1 - \cos \phi \quad (1.5.3)$$

where ϕ is the angle between the vectors \vec{k} and \vec{l} . When the two vectors have equal directions, the cosine will add to one. The angle ϕ can also be described as a function of \vec{k} and \vec{l} :

$$\cos \phi = \frac{\vec{k} \cdot \vec{l}}{\|\vec{k}\| \|\vec{l}\|} \quad (1.5.4)$$

The cosine distance is well suited for features that are real vectors and not a collection of independent scalar features.

The histogram intersection distance compares two histograms \vec{k} and \vec{l} of n bins by taking the intersection of both histograms:

$$\mathcal{D}_H(\vec{k}, \vec{l}) = 1 - \frac{\sum_{i=1}^n \min(k_i, l_i)}{\sum_{i=1}^n k_i} \quad (1.5.5)$$

When considering images of different sizes, this distance function is not a metric due to $\mathcal{D}_H(\vec{k}, \vec{l}) \neq \mathcal{D}_H(\vec{l}, \vec{k})$. In order to become a valid distance metric, histograms need to be normalized first:

$$\vec{k}^n = \frac{\vec{k}}{\sum_i^n k_i} \quad (1.5.6)$$

For normalized histograms (total sum of 1), the histogram intersection is given by:

$$\mathcal{D}_H^n(\vec{k}^n, \vec{l}^n) = 1 - \sum_i^n |k_i^n - l_i^n| \quad (1.5.7)$$

This is again the Minkowski-form distance metric with $k = 1$. Histogram intersection has the property that it allows for occlusion, i.e. when an object in one image is partly occluded, the visible part still contributes to the similarity [60], [59].

Alternative, histogram matching is proposed defined by normalized cross correlation:

$$\mathcal{D}_x(\vec{k}, \vec{l}) = \frac{\sum_{i=1}^n k_i l_i}{\sum_{i=1}^n k_i^2} \quad (1.5.8)$$

The normalized cross correlation has a maximum of unity that occurs if and only if \vec{k} exactly matches \vec{l} .

In the QBIC system [42], the weighted Euclidean distance has been used for the similarity of color histograms. In fact, the distance measure is based on the correlation between histograms \vec{k} and \vec{l} :

$$\mathcal{D}_Q(\vec{k}, \vec{l}) = (k_i - l_i)^t A (k_i - l_j) \quad (1.5.9)$$

Further, A is a weight matrix with term a_{ij} expressing the perceptual distance between bin i and j .

The average color distance has been proposed by [70] to obtain a simpler low-dimensional distance measure:

$$\mathcal{D}_{\text{Haf}}(\vec{k}, \vec{l}) = (k_{\text{avg}} - l_{\text{avg}})^t (k_{\text{avg}} - l_{\text{avg}}) \quad (1.5.10)$$

where k_{avg} and l_{avg} are 3x1 average color vectors of \vec{k} and \vec{l} .

As stated before, for broad domains, a proper similarity measure should be robust to object fragmentation, occlusion and clutter by the presence of other objects in the view. In [58], various similarity functions were compared for color-based histogram matching. From these results, it is concluded that retrieval accuracy of similarity functions depend on the presence of object clutter in the scene. The histogram cross correlation provide best retrieval accuracy without any object clutter (narrow domain). This is due to the fact that this similarity functions is symmetric and can be interpreted as the number of pixels with the same values in the query image which can be found present in the retrieved image and vice versa. This is a desirable property when one object per image is recorded without any object clutter. In the presence of object clutter (broad domain), highest image retrieval accuracy is provided by the quadratic similarity function (e.g. histogram intersection). This is because this similarity measure count the number of similar hits and hence are insensitive to false positives.

Finally, the natural measure to compare ordered sets of accumulative features is non-parametric test statistics. They can be applied to the distributions of the coefficients of transforms to determine the likelihood that two samples derive from the same distribution [14], [131]. They can also be applied to compare the equality of two histograms and all variations thereof.

1.5.3 Similarity of object outlines

In [176] a good review is given of methods to compare shapes directly after segmentation into a set of object points $t(\mathbf{x})$ without an intermediate description in terms of shape features.

For shape comparison, the authors make a distinction between transforms, moments, deformation matching, scale space matching and dissimilarity measurement. Difficulties for shape matching based on global transforms are the inexplicability of the result, and the brittleness for small deviations. Moments, specifically their invariant combinations, have been frequently used in retrieval [94]. Matching a query and an object in the data file can be done along the ordered set of eigen shapes [145], or with elastic matching [34], [11]. Scale space matching is based on progressively simplifying the contour by smoothing [118]. By comparing the signature of annihilated zero crossings of the curvature, two shapes are matched in a scale and rotation invariant fashion. A discrete analogue can be found in [98] where points are removed from the digitized contour on the basis of perceptually motivated rules.

When based on a metric, dissimilarity measures will render an ordered range of deviations, suited for a predictable interpretation. In [176], an analysis is given for the Hausdorff and related metrics between two shapes on robustness and computational complexity. The directed Hausdorff metric is defined as the maximum distance between a point on query object q and its closest counterpart on d . The partial Hausdorff metric, defined as the k -th maximum rather than the absolute maximum, is used in [71] for affine invariant retrieval.

1.5.4 Similarity of object arrangements

The result of a structural description is a hierarchically ordered set of feature values H . In this section we consider the similarity of two structural or layout descriptions.

Many different techniques have been reported for the similarity of feature structures. In [180], [82] a Bayesian framework is developed for the matching of relational attributed graphs by discrete relaxation. This is applied to line patterns from aerial photographs.

A metric for the comparison of two topological arrangements of named parts, applied to medical images, is defined in [166]. The distance is derived from the number of edit-steps needed to nullify the difference in the Voronoi-diagrams of two images.

In [18], 2D-strings describing spatial relationships between objects are discussed, and much later reviewed in [185]. From such topological relationships of image regions, in [79] a 2D-indexing is built in trees of symbol strings each representing the projection of a region on the co-ordinate axis. The distance between the H_q and H_d is the weighed number of editing operations required to transform the one tree to the other. In [151], a graph is formed from the image on the basis of symmetry as appears from the medial axis. Similarity is assessed in two stages via graph-based matching, followed by energy-deformation matching.

In [51], hierarchically ordered trees are compared for the purpose of retrieval by rewriting them into strings. A distance-based similarity measure establishes the similarity scores between corresponding leaves in the trees. At the level of trees, the total similarity score of corresponding branches is taken as the measure for (sub)tree-similarity. From a small size experiment, it is concluded that hierarchically ordered feature sets are more efficient than plain feature sets, with projected computational shortcuts for larger data sets.

1.5.5 Similarity of salient features

Salient features are used to capture the information in the image in a limited number of salient points. Similarity between images can then be checked in several different ways.

In the first place, the color, texture or local shape characteristics may be used to identify the salient points of the data as identical to the salient points of the query.

A measure of similarity between the feature values measured of the blobs resulting from weak segmentation consists of a Mahalanobis distance between the feature vector composed of the color, texture, position, area, eccentricity, and direction of the two ellipses [16].

In the second place, one can store all salient points from one image in a histogram on the basis of a few characteristics, such as color on the inside versus color on the outside. The similarity is then based on the group-wise presence of enough similar points [59]. The intersection model has been used in image retrieval in [153], while keeping access to their location in the image by back-projection [162]. Further, a weight per dimension may favor the appearance of some salient features over

another. See also [77] for a comparison with correlograms.

A third alternative for similarity of salient points is to concentrate only on the spatial relationships among the salient points sets. In point by point based methods for shape comparison, shape similarity is studied in [89], where maximum curvature points on the contour and the length between them are used to characterize the object. To avoid the extensive computations, one can compute the algebraic invariants of point sets, known as the cross-ratio. Due to their invariant character, these measures tend to have only a limited discriminatory power among different objects. A more recent version for the similarity of nameless point-sets is found in geometric hashing [182] where each triplet spans a base for the remaining points of the object. An unknown object is compared on each triplet to see whether enough similarly located points are found. Geometric hashing, though attractive in its concept, is too computationally expensive to be used on the very large data sets of image retrieval due to the anonymity of the points. Similarity of two points sets given in a row-wise matrix is discussed in [179].

1.5.6 Discussion

Whenever the image itself permits an obvious interpretation, the ideal content-based system should employ that information. A strong semantic interpretation occurs when a sign can be positively identified in the image. This is rarely the case due to the large variety of signs in a broad class of images and the enormity of the task to define a reliable detection algorithm for each of them. Weak semantics rely on inexact categorization induced by similarity measures, preferably online by interaction. The categorization may agree with semantic concepts of the user, but the agreement is in general imperfect. Therefore, the use of weak semantics is usually paired with the ability to gear the semantics of the user to his or her needs by interpretation. Tunable semantics is likely to receive more attention in the future especially when data sets grow big.

Similarity is an interpretation of the image based on the difference with another image. For each of the feature types a different similarity measure is needed. For similarity between feature sets, special attention has gone to establishing similarity among histograms due to their computational efficiency and retrieval effectiveness.

Similarity of shape has received a considerable attention in the context of object-based retrieval. Generally, global shape matching schemes break down when there is occlusion or clutter in the scene. Most global shape comparison methods implicitly require a frontal viewpoint against a clear enough background to achieve a sufficiently precise segmentation. With the recent inclusion of perceptually robust points in the shape of objects, an important step forward has been made.

Similarity of hierarchically ordered descriptions deserves considerable attention, as it is one mechanism to circumvent the problems with segmentation while maintaining some of the semantically meaningful relationships in the image. Part of the difficulty here is to provide matching of partial disturbances in the hierarchical order and the influence of sensor-related variances in the description.

1.6 Interaction and Learning

1.6.1 Interaction on a semantic level

In [78], knowledge-based type abstraction hierarchies are used to access image data based on context and a user profile, generated automatically from cluster analysis of the database. Also in [19], the aim is to create a very large concept-space inspired by the thesaurus-based search from the information retrieval community. In [117] a linguistic description of texture patch visual qualities is given, and ordered in a hierarchy of perceptual importance on the basis of extensive psychological experimentation.

A more general concept of similarity is needed for relevance feedback, in which similarity with respect to an ensemble of images is required. To that end, in [43] more complex relationships are presented between similarity and distance functions defining a weighted measure of two simpler similarities $S(s, S_1, S_2) = w_1 \exp(-d(S_1, s)) + w_2 \exp(-d(S_2, s))$. The purpose of the bi-referential measure is to find all regions that are similar to two specified query points, an idea that generalizes to similarity queries given multiple examples. The approach can be extended with the definition of a complete algebra of similarity measures with suitable composition operators [43], [38]. It is then possible to define operators corresponding to the disjunction, conjunction, and negation of similarity measures, much like traditional databases. The algebra is useful for the user to manipulate the similarity directly as a means to express characteristics in specific feature values.

1.6.2 Classification on a semantic level

To further enhance the performance of content-based retrieval systems, image classification has been proposed to group images into semantically meaningful classes [171], [172], [184], [188]. The advantage of these classification schemes is that simple, low-level image features can be used to express semantically meaningful classes. Image classification is based on unsupervised learning techniques such as clustering, Self-Organization Maps (SOM) [188] and Markov models [184]. Further, supervised grouping can be applied. For example, vacation images have been classified based on a Bayesian framework into city vs. landscape by supervised learning techniques [171], [172]. However, these classification schemes are entirely based on pictorial information. Aside from image *retrieval* ([44], [146]), very little attention has been paid on using both textual and pictorial information for *classifying* images on the Web. This is even more surprisingly if one realizes that images on Web pages are usually surrounded by text and discriminatory HTML tags such as IMG, and the HTML fields SRC and ALT. Hence, WWW images have intrinsic annotation information induced by the HTML structure. Consequently, the set of images on the Web can be seen as an annotated image set.

1.6.3 Learning

As data sets grow big and the processing power matches that growth, the opportunity arises to learn from experience. Rather than designing, implementing and testing an algorithm to detect the visual characteristics for each different semantic term, the aim is to learn from the appearance of objects directly.

For a review on statistical pattern recognition, see [2]. In [174] a variety of techniques is discussed treating retrieval as a classification problem.

One approach is principal component analysis over a stack of images taken from the same class z of objects. This can be done in feature space [120] or at the level of the entire image, for examples faces in [115]. The analysis yields a set of eigenface images, capturing the common characteristics of a face without having a geometric model.

Effective ways to learn from partially labeled data have recently been introduced in [183], [32] both using the principle of transduction [173]. This saves the effort of labeling the entire data set, infeasible and unreliable as it grows big.

In [169] a very large number of pre-computed features is considered, of which a small subset is selected by boosting [2] to learn the image class.

An interesting technique to bridge the gap between textual and pictorial descriptions to exploit information at the level of documents is borrowed from information retrieval, called latent semantic indexing [146], [187]. First a corpus is formed of documents (in this case images with a caption) from which features are computed. Then by singular value decomposition, the dictionary covering the captions is correlated with the features derived from the pictures. The search is for hidden correlations of features and captions.

1.6.4 Discussion

Learning computational models for semantics is an interesting and relatively new approach. It gains attention quickly as the data sets and the machine power grow big. Learning opens up the possibility to an interpretation of the image without designing and testing a detector for each new notion. One such approach is appearance-based learning of the common characteristics of stacks of images from the same class. Appearance-based learning is suited for narrow domains. For the success of the learning approach there is a trade-off between standardizing the objects in the data set and the size of the data set. The more standardized the data are the less data will be needed, but, on the other hand, the less broadly applicable the result will be. Interesting approaches to derive semantic classes from captions, or a partially labeled or unlabeled data set have been presented recently, see above.

1.7 Conclusion

In this chapter, we have presented an overview on the theory, techniques and applications of content-based image retrieval. We took patterns of use and computation as the pivotal building blocks of our survey.

From a scientific perspective the following trends can be distinguished. First, large scale image databases are being created. Obviously, large scale datasets provide different image mining problems than rather small, narrow- domain datasets. Second, research is directed towards the integration of different information modalities such as text, pictorial, and motion. Third, relevance feedback will be and still is an important issue. Finally, invariance is necessary to get to general-purpose image retrieval.

From a societal/commercial perspective, it is obvious that there will be enormous increase in the amount of digital images used in various communication frameworks such as promotion, sports, education, and publishing. Further, digital images have become one of the major multimedia information sources on Internet, where the amount of image/video on the Web is growing each day. Moreover, with the introduction of the new generation cell-phones, a tremendous market will be opened for the storage and management of pictorial data. Due to this tremendous amount of pictorial information, image mining and search tools are required as indexing, searching and assessing the content of large scale image databases is inherently a time-consuming operation when done by human operators. Therefore, product suites for content-based video indexing and searching is not only necessary but essential for future content owners in the field of entertainment, news, education, video communication and distribution.

We hope that from this review that you get the picture in this new pictorial world...

BIBLIOGRAPHY

- [1] R.K. Srihari A. Rao and Z. Zhang. Geometric histogram: A distribution of geometric configurations of color subsets. In *Internet Imaging*, volume 3,964, pages 91–101, 2000.
- [2] R.P.W. Duin A.K. Jain and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on PAMI*, 22(1):4 – 37, 2000.
- [3] R. Alferez and Y-F Wang. Geometric and illumination invariants for object recognition. *IEEE Transactions on PAMI*, 21(6):505 – 536, 1999.
- [4] D. Androustos, K. N. Plataniotis, and A. N. Venetsanopoulos. A novel vector-based approach to color image retrieval using a vector angular-based distance measure. *Image Understanding*, 75(1-2):46–58, 1999.
- [5] E. Angelopoulou and L. B. Wolff. Sign of gaussian curvature from curve orientation in photometric space. *IEEE Transactions on PAMI*, 20(10):1056 – 1066, 1998.
- [6] A.R. Appas, A.M. Darwish, A.I. El-Desouki, and S.I. Shaheen. Image indexing using composite regional color channel features. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases VII*, pages 492–500, 1999.
- [7] L. Armitage and P. Enser. Analysis of user need in image archives. *Journal of Information Science*, 23(4):287–299, 1997.
- [8] F.G. Ashby and N. A. Perrin. Toward a unified theory of similarity and recognition. *Psychological Review*, 95(1):124–150, 1988.
- [9] D. Ashlock and J. Davidson. Texture synthesis with tandem genetic algorithms using nonparametric partially ordered markov models. In *Proceedings of the Congress on Evolutionary Computation (CEC99)*, pages 1157–1163, 1999.
- [10] R. Bajcsy, S. W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by inter-reflections. In *IEEE 10th ICPR'90*, pages 785–790, Atlantic City, NJ, 1990.

- [11] R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38(15-16):2365–2385, 1998.
- [12] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using em and its application to content-based image retrieval. In *Sixth International Conference on Computer Vision*, 1998.
- [13] A. Del Bimbo, M. Mugnaini, P. Pala, and F. Turco. Visual querying by color perceptive regions. *Pattern Recognition*, 31(9):1241–1253, 1998.
- [14] J. De Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Computer Vision and Pattern Recognition*, 1998.
- [15] 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.
- [16] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *Proceedings of the IEEE International Workshop on Content-Based Access of Image and Video Databases*, 1997.
- [17] S. K. Chang and A. D. Hsu. Image-information systems - where do we go from here. *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431 – 442, 1992.
- [18] S. K. Chang, Q. Y. Shi, and C. W. Yan. Iconic indexing by 2d strings. *IEEE Transactions on PAMI*, 9:413 – 428, 1987.
- [19] H. Chen, B. Schatz, T. Ng, J. Martinez, A. Kirchhoff, and C. Lim. A parallel computing approach to creating engineering concept spaces for semantic retrieval: the Illinois digital library initiative project. *IEEE Transactions on PAMI*, 18(8):771 – 782, 1996.
- [20] Y. Chen and E.K. Wong. Augmented image histogram for image and video similarity search. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases VII*, pages 523–429, 1999.
- [21] H. Choi and R. Baraniuk. Multiscale texture segmentation using wavelet-domain hidden markov models. In *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1692–1697, 1998.
- [22] C. K. Chui, L. Montefusco, and L. Puccio. *Wavelets : theory, algorithms, and applications*. Academic Press, San Diego, 1994.
- [23] L. Cinque, S. Levialdi, and A. Pellicano. Color-based image retrieval using spatial-chromatic histograms. In *IEEE Multimedia Systems*, volume 2, pages 969–973, 1999.
- [24] G. Ciocca and R. Schettini. A relevance feedback mechanism for content-based image retrieval. *Information Processing and Management*, 35:605–632, 1999.

- [25] G. Ciocca and R. Schettini. Using a relevance feedback mechanism to improve content-based image retrieval. In *Proceedings of Visual Information and Information Systems*, pages 107–114, 1999.
- [26] C. Colombo, A. Rizzi, and I. Genovesi. Histogram families for color-based retrieval in image databases. In *Proc. ICIAP'97*, 1997.
- [27] P. Correia and F. Pereira. The role of analysis in content-based video coding and indexing. *Signal Processing*, 66(2):125 – 142, 1998.
- [28] J.M. Corridoni, A. del Bimbo, and P. Pala. Image retrieval by color semantics. *Multimedia systems*, 7:175 – 183, 1999.
- [29] I. J. Cox, M. L. Miller, T. P. Minka, and T. V. Pappas. The bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE Transactions on Image Processing*, 9(1):20 – 37, 2000.
- [30] G. Csurka and O. Faugeras. Algebraic and geometrical tools to compute projective and permutation invariants. *IEEE Transactions on PAMI*, 21(1):58 – 65, 1999.
- [31] J. F. Cullen, J. J. Hull, and P. E. Hart. Document image database retrieval and browsing using texture analysis. In *Proceedings of the fourth international conference on document analysis and recognition, Ulm, Germany*, pages 718–721, 1997.
- [32] M.-H. Yang D. Roth and N. Ahuja. Learning to recognize objects. In *Computer Vision and Pattern Recognition*, pages 724–731, 2000.
- [33] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [34] A. del Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on PAMI*, 19(2):121–132, 1997.
- [35] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.
- [36] E. Di Sciascio Di, G. Mingolla, and M. Mongiello. Content-based image retrieval over the web using query by sketch and relevance feedback. In *VISUAL99*, pages 123–30, 1999.
- [37] A. Dimai. Spatial encoding using differences of global features. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases IV*, pages 352–360, 1997.
- [38] D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Information Sciences*, 36:85–121, 1985.

- [39] J.P. Eakins, J.M. Boardman, and M.E. Graham. Similarity retrieval of trademark images. *IEEE Multimedia*, 5(2):53–63, 1998.
- [40] C. Esperanca and H. Samet. A differential code for shape representation in image database applications. In *Proceedings of the IEEE International Conference on Image Processing Santa Barbara, CA, USA, 1997*.
- [41] L.M. Kaplan et. al. Fast texture database retrieval using extended fractal features. In I. Sethi and R. Jain, editors, *Proceedings of SPIE vol. 3312, Storage and Retrieval for Image and Video Databases, VI*, pages 162–173, 1998.
- [42] M. Flicker et al. Query by image and video content: the qbic system. *IEEE Computer*, 28(9), 1995.
- [43] R. Fagin. Combining fuzzy information from multiple systems. *J Comput Syst Sci*, 58(1):83–99, 1999.
- [44] J. Favella and V. Meza. Image-retrieval agent: Integrating image content and text. 1999.
- [45] G. D. Finlayson, S. S. Chatterjee, and B. V. Funt. Color angular indexing. In *ECCV96*, pages 16–27, 1996.
- [46] G.D. Finlayson, M.S. Drew, and B.V. Funt. Spectral sharpening: Sensor transformation for improved color constancy. *JOSA*, 11:1553–1563, 1994.
- [47] 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*, 1995.
- [48] D. Forsyth. Novel algorithm for color constancy. *International Journal of Computer Vision*, 5:5–36, 1990.
- [49] D.A. Forsyth and M.M. Fleck. Automatic detection of human nudes. *International Journal of Computer Vision*, 32(1):63–77, 1999.
- [50] G. Frederix and E.J. Pauwels. Automatic interpretation based on robust segmentation and shape extraction. In D.P. Huijsmans and A.W.M. Smeulders, editors, *Proceedings of Visual 99, International Conference on Visual Information Systems*, volume 1614 of *Lecture Notes in Computer Science*, pages 769–776, 1999.
- [51] C-S. Fuh, S-W Cho, and K. Essig. Hierarchical color image region segmentation for content-based image retrieval system. *IEEE Transactions on Image Processing*, 9(1):156 – 163, 2000.
- [52] B.V. Funt and M.S. Drew. Color constancy computation in near-mondrian scenes. In *Computer Vision and Pattern Recognition*, pages 544–549, 1988.

-
- [53] B.V. Funt and G.D. Finlayson. Color constant color indexing. *IEEE Transactions on PAMI*, 17(5):522–529, 1995.
- [54] J. M. Geusebroek, A. W. M. Smeulders, and R. van den Boomgaard. Measurement of color invariants. In *Computer Vision and Pattern Recognition*. IEEE Press, 2000.
- [55] Th. Gevers. Color based image retrieval. In *Multimedia Search*. Springer Verlag, 2000.
- [56] Th. Gevers. Image segmentation and matching of color-texture objects. *IEEE Trans. on Multimedia*, 4(4), 2002.
- [57] Th. Gevers and A. W. M. Smeulders. Color based object recognition. *Pattern recognition*, 32(3):453 – 464, 1999.
- [58] Th. Gevers and A. W. M. Smeulders. Content-based image retrieval by viewpoint-invariant image indexing. *Image and Vision Computing*, 17(7):475 – 488, 1999.
- [59] Th. Gevers and A. W. M. Smeulders. Pictoseek: combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing*, 9(1):102 – 119, 2000.
- [60] Th. Gevers and A.W.M. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, 1999.
- [61] Th. Gevers and H. M. G. Stokman. Classification of color edges in video into shadow-geometry, highlight, or material transitions. *IEEE Trans. on Multimedia*, 5(2), 2003.
- [62] Th. Gevers and H. M. G. Stokman. Robust histogram construction from color invariants for object recognition. *IEEE Transactions on PAMI*, 25(10), 2003.
- [63] Th. Gevers, P. Vreman, and J. van der Weijer. Color constant texture segmentation. In *IS&T/SPIE Symposium on Electronic Imaging: Internet Imaging I*, 2000.
- [64] G.L. Gimel’farb and A. K. Jain. On retrieving textured images from an image database. *Pattern Recognition*, 29(9):1461–1483, 1996.
- [65] Y. Gong, C.H. Chuan, and G. Xiaoyi. Image indexing and retrieval using color histograms. *Multimedia Tools and Applications*, 2:133–156, 1996.
- [66] C. C. Gottlieb and H. E. Kreyszig. Texture descriptors based on co-occurrences matrices. *Computer Vision, Graphics, and Image Processing*, 51, 1990.
- [67] L.J. Guibas, B. Rogoff, and C. Tomasi. Fixed-window image descriptors for image retrieval. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases III*, pages 352–362, 1995.

- [68] A. Gupta and R. Jain. Visual information retrieval. *Communications of the ACM*, 40(5):71–79, 1997.
- [69] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *ACM SIGMOD*, pages 47 – 57, 1984.
- [70] J. Hafner, H.S. Sawhney, W. Equit, M. Flickner, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on PAMI*, 17(7):729–736, 1995.
- [71] M. Hagendoorn and R. C. Veltkamp. Reliable and efficient pattern matching using an affine invariant metric. *International Journal of Computer Vision*, 35(3):203 – 225, 1999.
- [72] S. Hastings. Query categories in a study of intellectual access to digitized art images. In *ASIS '95, Proceedings of the 58th Annual Meeting of the American Society for Information Science, Chicago, IL*, 1995.
- [73] G. Healey. Segmenting images using normalized color. *IEEE Transactions on Systems, Man and Cybernetics*, 22(1):64–73, 1992.
- [74] G. Healey and D. Slater. Computing illumination-invariant descriptors of spatially filtered color image regions. *IEEE Transactions on Image Processing*, 6(7):1002 – 1013, 1997.
- [75] K. Hirata and T. Kato. Rough sketch-based image information retrieval. *NEC Res Dev*, 34(2):263 – 273, 1992.
- [76] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori. Visualization of information spaces to retrieve and browse image data. In D.P. Huijsmans and A.W.M. Smeulders, editors, *Proceedings of Visual 99, International Conference on Visual Information Systems*, volume 1614 of *Lecture Notes in Computer Science*, pages 155–162, 1999.
- [77] N.R. Howe and D.P. Huttenlocher. Integrating color, texture, and geometry for image retrieval. In *Computer Vision and Pattern Recognition*, pages 239–247, 2000.
- [78] C.C. Hsu, W.W. Chu, and R.K. Taira. A knowledge-based approach for retrieving images by content. *IEEE Transactions on Knowledge and Data Engineering*, 8(4):522–532, 1996.
- [79] F. J. Hsu, S. Y. Lee, and B. S. Lin. Similarity retrieval by 2D C-trees matching in image databases. *Journal of Visual Communication and Image Representation*, 9(1):87 – 100, 1998.
- [80] J. Huang, S. R. Kumar, M. Mitra, W-J Zhu, and R. Zabih. Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3):245 – 268, 1999.

- [81] J. Huang, S.R. Kumar, M. Mitra, W-J Zhu, and R. Ramin. Image indexing using color correlograms. In *Computer Vision and Pattern Recognition*, pages 762–768, 1997.
- [82] B. Huet and E. R. Hancock. Line pattern retrieval using relational histograms. *IEEE Transactions on PAMI*, 21(12):1363 – 1371, 1999.
- [83] F. Idris and S. Panchanathan. Image indexing using wavelet vector quantization. In *Proceedings of the SPIE Vol. 2606-Digital Image Storage and Archiving Systems*, pages 269–275, 1995.
- [84] L. Itti, C. Koch, and E. Niebur. A model for saliency-based visual attention for rapid scene analysis. *IEEE Transactions on PAMI*, 20(11):1254 – 1259, 1998.
- [85] C.E. Jacobs, A. Finkelstein, and D.H. Salesin. Fast multiresolution image querying. In *Computer Graphics*, 1995.
- [86] C.E. Jacobs, A. Finkelstein, and S. H. Salesin. Fast multiresolution image querying. In *Proceedings of SIGGRAPH 95, Los Angeles, CA*. ACM SIGGRAPH, New York, 1995.
- [87] A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.
- [88] A. K. Jain and A. Vailaya. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9):1369 – 1390, 1998.
- [89] L. Jia and L. Kitchen. Object-based image similarity computation using inductive learning of contour-segment relations. *IEEE Transactions on Image Processing*, 9(1):80 – 87, 2000.
- [90] D. W. Joyce, P. H. Lewis, R. H. Tansley, M. R. Dobie, and W. Hall. Semiotics and agents for integrating and navigating through multimedia representations. In Minerva M. Yeung, Boon-Lock Yeo, and Charles Bouman, editors, *Proceedings of SPIE Vol. 3972, Storage and Retrieval for Media Databases 2000*, pages 120–131, 2000.
- [91] T. Kato, T. Kurita, N. Otsu, and K. Hirata. A sketch retrieval method for full color image database - query by visual example. In *Proceedings of the ICPR, Computer Vision and Applications, The Hague*, pages 530–533, 1992.
- [92] J.R. Kender. Saturation, hue, and normalized colors: Calculation, digitization effects, and use. Technical report, Department of Computer Science, Carnegie-Mellon University, 1976.
- [93] G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal Computer Vision*, pages 7–38, 4 1990.

- [94] A. Kontanzad and Y. H. Hong. Invariant image recognition by Zernike moments. *IEEE Transactions on PAMI*, 12(5):489 – 497, 1990.
- [95] S. Krishnamachari and R. Chellappa. Multiresolution gauss-markov random field models for texture segmentation. *IEEE Transactions on Image Processing*, 6(2), 1997.
- [96] A. Laine and J. Fan. Texture classification by wavelet packet signature. *IEEE Transactions on PAMI*, 15(11):1186–1191, 1993.
- [97] E. H. Land. The retinex theory of color vision. *Scientific American*, 218(6):108–128, 1977.
- [98] L. J. Latecki and R. Lakamper. Convexity rule for shape decomposition based on discrete contour evolution. *Image Understanding*, 73(3):441 – 454, 1999.
- [99] K-S. Leung and R. Ng. Multiresolution subimage similarity matching for large image databases. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases VI*, pages 259–270, 1998.
- [100] C-S. Li and V. Castelli. Deriving texture feature set for content-based retrieval of satellite image database. In *Proceedings of the IEEE International Conference on Image Processing Santa Barbara, CA, USA, 1997*.
- [101] K.C. Liang and C.C.J. Kuo. Progressive image indexing and retrieval based on embedded wavelet coding. In *IEEE International Conference on Image Processing*, volume 1, pages 572–575, 1997.
- [102] H. C. Lin, L. L. Wang, and S. N. Yang. Color image retrieval based on hidden Markov models. *IEEE Transactions on Image Processing*, 6(2):332 – 339, 1997.
- [103] T. Lindeberg and J. O. Eklundh. Scale space primal sketch construction and experiments. *Journ Image Vis Comp*, 10:3 – 18, 1992.
- [104] F. Liu and R. Picard. Periodicity, directionality, and randomness: Wold features for image modelling and retrieval. *IEEE Transactions on PAMI*, 18(7):517–549, 1996.
- [105] M. Welling M. Weber and P. Perona. Towards automatic discovery of object categories. In *Computer Vision and Pattern Recognition*, pages 101–108, 2000.
- [106] W. Y. Ma and B. S. Manjunath. Edge flow: a framework of boundary detection and image segmentation. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico*, pages 744–749, 1997.
- [107] J. Malki, N. Boujemaa, C. Nastar, and A. Winter. Region queries without segmentation for image retrieval content. In *Int. Conf. on Visual Information Systems, VISUAL99*, pages 115–122, 1999.

- [108] M. K. Mandal, F. Idris, and S. Panchanathan. Image and video indexing in the compressed domain: a critical review. *Image and Vision Computing*, 2000.
- [109] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on PAMI*, 18(8):837 – 842, 1996.
- [110] J. Mao and A.K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2), 1992.
- [111] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. In *Proc. 5th ICCV*, pages 726 – 732, 1995.
- [112] R. Mehrotra and J. E. Gary. Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9):57–62, 1995.
- [113] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee. Shape measures for content based image retrieval: A comparison. *Information Proc. Management*, 33(3):319 – 337, 1997.
- [114] M. Mirmehdi and M. Petrou. Segmentation of color texture. *PAMI*, 22(2):142 – 159, 2000.
- [115] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on PAMI*, 19(7):696 – 710, 1997.
- [116] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. In *3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan*, 1998.
- [117] A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, and S. K. Ganapathy. Matching and retrieval based on the vocabulary and grammar of color patterns. *IEEE Transactions on Image Processing*, 9(1):38 – 54, 2000.
- [118] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale-space. *IEEE Transactions on PAMI*, 17(5):539–544, 1995.
- [119] J. L. Mundy, A. Zissermann, and D. Forsyth, editors. *Applications of invariance in computer vision*, volume 825 of *Lecture Notes in Computer Science*. Springer Verlag GmbH, 1994.
- [120] H. Murase and S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14(1):5 – 24, 1995.
- [121] S. K. Nayar and R. M. Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 17(3):219–240, 1996.
- [122] T. Ojala, M. Pietikainen, and D. Harwood. A comparison study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51 – 59, 1996.

- [123] P. Pala and S. Santini. Image retrieval by shape and texture. *Pattern Recognition*, 32(3):517–527, 1999.
- [124] D.K. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on PAMI*, 17(10):939 – 954, 1995.
- [125] G. Pass and R. Zabith. Comparing images using joint histograms. *Multimedia systems*, 7:234 – 240, 1999.
- [126] E. J. Pauwels and G. Frederix. Nonparametric clustering for image segmentation and grouping. *Image Understanding*, 75(1):73 – 85, 2000.
- [127] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233 – 254, 1996.
- [128] E. Petrakis and C. Faloutsos. Similarity searching in medical image databases. *IEEE Transactions on Knowledge and Data Engineering*, 9(3):435–447, 1997.
- [129] R.W. Picard and T.P. Minka. Vision texture for annotation.
- [130] M. Pietikainen, S. Nieminen, E. Marszalec, and T. Ojala. Accurate color discrimination with classification based on feature distributions. In *Proc. Int'l Conf. Pattern Recognition*, pages 833 – 838, 1996.
- [131] J. Puzicha, T. Hoffman, and J. M. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition-CVPR*, 1997.
- [132] W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, D. Venugopal, D. S. Song, and R. A. Clark. Tree-structured wavelet transform segmentation of microcalcifications in digital mammography. *Jl Med. Phys.*, 22(8):1247 – 1254, 1995.
- [133] T. Randen and J. Hakon Husoy. Filtering for texture classification: a comparative study. *IEEE Transactions on PAMI*, 21(4):291 – 310, 1999.
- [134] E. Riloff and L. Hollaar. Text databases and information retrieval. *ACM Computing Surveys*, 28(1):133–135, 1996.
- [135] E. Rivlin and I. Weiss. Local invariants for recognition. *IEEE Transactions on PAMI*, 17(3):226 – 238, 1995.
- [136] R. Rodriguez-Sanchez, J. A. Garcia, J. Fdez-Valdivia, and X. R. Fdez-Vidal. The rgff representational model: a system for the automatically learned partitioning of 'visual pattern' in digital images. *IEEE Transactions on PAMI*, 21(10):1044 – 1073, 1999.

- [137] P. L. Rosin. Edges: Saliency measures and automatic thresholding. *Machine Vision and Appl*, 9(7):139 – 159, 1997.
- [138] I. Rothe, H. Suesse, and K. Voss. The method of normalization of determine invariants. *IEEE Transactions on PAMI*, 18(4):366 – 376, 1996.
- [139] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on circuits and video technology*, 1998.
- [140] M. Beigi S.-F. Chang, J.R. Smith and A. Benitez. Visual information retrieval from large distributed online repositories. *Comm. ACM*, 40(12):63 – 71, 1997.
- [141] S. Santini, A. Gupta, and R. Jain. User interfaces for emergent semantics in image databases. In *Proceedings of the 8th IFIP Working Conference on Database Semantics (DS-8), Rotorua (New Zealand)*, 1999.
- [142] S. Santini and R. Jain. Similarity measures. *IEEE Transactions on PAMI*, 21(9):871 – 883, 1999.
- [143] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on PAMI*, 19(5):530 – 535, 1997.
- [144] M. Schneier and M. Abdel-Mottaleb. Exploiting the JPEG compression scheme for image retrieval. *IEEE Transactions on PAMI*, 18(8):849 – 853, 1996.
- [145] S. Sclaroff. Deformable prototypes for encoding shape categories in image databases. *Pattern Recognition*, 30(4):627 – 641, 1997.
- [146] S. Sclaroff, M. LaCascia, and S. Sethi. Using textual and visual cues for content-based image retrieval from the World Wide Web. *Image Understanding*, 75(2):86 – 98, 1999.
- [147] S. Sclaroff, L. Taycher, and M. La Cascia. Imagerover: A content-based image browser for the world wide web. In *IEEE Workshop on Content-based Access and Video Libraries*, 1997.
- [148] N. Sebe, M.S. Lew, and D.P. Huijsmands. Multi-scale sub-image search. In *ACM Int. Conf. on Multimedia*, 1999.
- [149] S. Servetto, Y. Rui, K. Ramchandran, and T. S. Huang. A region-based representation of images in mars. *Journal on VLSI Signal Processing Systems*, 20(2):137–150, 1998.
- [150] S.A. Shafer. Using color to separate reflection components. *COLOR Res. Appl.*, 10(4):210–218, 1985.
- [151] D. Sharvit, J. Chan, H. Tek, and B. B. Kimia. Symmetry-based indexing of image databases. *Journal of Visual Communication and Image Representation*, 9(4):366 – 380, 1998.

- [152] R. N. Shepard. Toward a universal law of generalization for physical science. *Science*, 237:1317–1323, 1987.
- [153] D. Slater and G. Healey. The illumination-invariant recognition of 3D objects using local color invariants. *IEEE Transactions on PAMI*, 18(2):206 – 210, 1996.
- [154] A.W.M. Smeulders, M. L. Kersten, and Th. Gevers. Crossing the divide between computer vision and data bases in search of image databases. In *Fourth Working Conference on Visual Database Systems, L'Aquila, Italy*, pages 223–239, 1998.
- [155] 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 PAMI*, 22(12):1349 – 1380, 2000.
- [156] J. R. Smith and S. F. Chang. Automated binary feature sets for image retrieval. In C. Faloutsos, editor, *Proceedings of ICASSP, Atlanta*. Kluwer Academic, 1996.
- [157] J. R. Smith and S-F. Chang. Integrated spatial and feature image query. *Multimedia systems*, 7(2):129 – 140, 1999.
- [158] J.R. Smith and S-F. Chang. Visualseek: a fully automated content-based image query system. In *ACM Multimedia*, 1996.
- [159] S. M. Smith and J. M. Brady. SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45 – 78, 1997.
- [160] M. Stricker and M. Swain. The capacity of color histogram indexing. In *Computer Vision and Pattern Recognition*, pages 704 – 708. IEEE Press, 1994.
- [161] M.A. Stricker and M. Orengo. Similarity of color images. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases IV*, 1996.
- [162] M. J. Swain and B. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11 – 32, 1991.
- [163] M.J. Swain. Searching for multimedia on the world wide web. In *IEEE International Conference on Multimedia Computing and Systems*, pages 33–37, 1999.
- [164] D. J. Swets and J. Weng. Hierarchical discriminant analysis for image retrieval. *IEEE Transactions on PAMI*, 21(5):386 – 401, 1999.
- [165] T.F. Syeda-Mahmood. Data and model-driven selection using color regions. *International Journal of Computer Vision*, 21(1):9–36, 1997.
- [166] H. D. Tagare, F. M. Vos, C. C. Jaffe, and J. S. Duncan. Arrangement - a spatial relation between parts for evaluating similarity of tomographic section. *IEEE Transactions on PAMI*, 17(9):880 – 893, 1995.

- [167] T. Tan. Rotation invariant texture features and their use in automatic script identification. *IEEE Transactions on PAMI*, 20(7):751 – 756, 1998.
- [168] P. M. Tardif and A. Zaccarin. Multiscale autoregressive image representation for texture segmentation. In *Proceedings of SPIE Vol. 3026, Nonlinear Image Processing VIII, San Jose, CA, USA*, pages 327–337, 1997.
- [169] K. Tieu and P. Viola. Boosting image retrieval. In *Computer Vision and Pattern Recognition*, pages 228–235, 2000.
- [170] T. Tuytelaars and L. van Gool. Content-based image retrieval based on local affinely invariant regions. In *Proceedings of Visual Information and Information Systems*, pages 493 – 500, 1999.
- [171] A. Vailaya, M. Figueiredo, A. Jain, and H. Zhang. A bayesian framework for semantic classification of outdoor vacation images. In C. a. Bouman M. M. Yeung, B. Yeo, editor, *Storage and Retrieval for Image and Video Databases VII - SPIE*, pages 415–426, 1999.
- [172] A. Vailaya, M. Figueiredo, A. Jain, and H. Zhang. Content-based hierarchical classification of vacation images. In *IEEE International Conference on Multimedia Computing and Systems*, 1999.
- [173] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [174] N. Vasconcelos and A. Lippman. A probabilistic architecture for content-based image retrieval. In *Computer Vision and Pattern Recognition*, pages 216–221, 2000.
- [175] A. Vellaikal and C.C.J. Kuo. Content-based retrieval using multiresolution histogram representation. *Digital Image Storage Archiving Systems*, pages 312–323, 1995.
- [176] R. C. Veltkamp and M. Hagendoorn. State-of-the-art in shape matching. In *Multimedia search: state of the art*. Springer Verlag GmbH, 2000.
- [177] L. Z. Wang and G. Healey. Using Zernike moments for the illumination and geometry invariant classification of multispectral texture. *IEEE Transactions on Image Processing*, 7(2):196 – 203, 1991.
- [178] J. Weickert, S. Ishikawa, and A. Imiya. Linear scale space has first been proposed in japan. *Journal of Mathematical Imaging and Vision*, 10:237 – 252, 1999.
- [179] M. Werman and D. Weinshall. Similarity and affine invariant distances between 2d point sets. *IEEE Transactions on PAMI*, 17(8):810 – 814, 1995.
- [180] R. C. Wilson and E. R. Hancock. Structural matching by discrete relaxation. *IEEE Transactions on PAMI*, 19(6):634 – 648, 1997.

- [181] L. Wolff, S. A. Shafer, and G. E. Healey, editors. *Physics-based vision: principles and practice*, volume 2. Jones and Bartlett, Boston etc., 1992.
- [182] H.J. Wolfson and I. Rigoutsos. Geometric hashing: An overview. *IEEE computational science and engineering*, 4(4):10 – 21, 1997.
- [183] Q. Tian Y. Wu and T.S. Huang. Discriminant-em algorithm with applications to image retrieval. In *Computer Vision and Pattern Recognition*, pages 222–227, 2000.
- [184] H. H. Yu and W. Wolf. Scene classification methods for image and video databases. In *Proc. SPIE on Digital Image Storage and Archiving Systems*, pages 363–371, 1995.
- [185] Q. L. Zhang, S. K. Chang, and S. S. T. Yau. A unified approach to iconic indexing, retrieval and maintenance of spatial relationships in image databases. *Journal of Visual Communication and Image Representation*, 7(4):307 – 324, 1996.
- [186] Y.J. Zhang, Z.W. Liu, and Y. He. Comparison and improvement of color-based image retrieval. In *IS&T/SPIE Symposium on Electronic Imaging: Storage and Retrieval for Image and Video Databases IV*, pages 371–382, 1996.
- [187] R. Zhao and W. Grosky. Locating text in complex color images. In *IEEE International Conference on Multimedia Computing and Systems*, 2000.
- [188] Y. Zhong, K. Karu, and A. K. Jain. Locating text in complex color images. *Pattern Recognition*, 28(10):1523 – 1535, 1995.
- [189] P. Zhu and P. M. Chirlian. On critical point detection of digital shapes. *IEEE Transactions on PAMI*, 17(8):737 – 748, 1995.
- [190] S.C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on PAMI*, 18(9):884 – 900, 1996.