1 Visible Image Retrieval

Carlo Colombo Alberto Del Bimbo
Dipartimento di Sistemi e Informatica
Università di Firenze
Via Santa Marta 3,
I-50139 Firenze, ITALY
E-mail [colombo,delbimbo]@dsi.unifi.it

1.1 INTRODUCTION

The emergence of multimedia, the availability of large digital archives, as well as the rapid growth of the World-Wide Web, have recently attracted research efforts in providing tools for effective retrieval of image data based on their content (*Content-Based Image Retrieval*, CBIR). The relevance of CBIR for many applications, ranging from art galleries and museum archives, to picture/photograph, medical and geographic databases, criminal investigation, intellectual property and trademarks, fashion and interior design, makes this research field one of the fastest growing in information technology. Yet, after a decade of intensive research, CBIR technologies – save perhaps for very specialized areas such as crime prevention, medical diagnosis or fashion design – have had a limited impact on real-world applications. For instance, recent attempts to enhance text-based search engines on the WWW with CBIR options highlight both an increasing interest in the use of digital imagery and the current limitations of general-purpose image search facilities.

This chapter reviews applications and research themes in *Visible Image Retrieval*, namely, retrieval by content of heterogeneous collections of single images generated with visible spectrum technologies. It is generally agreed that a key design challenge in the field is how to reduce the semantic gap between user expectation and system support, especially in non-professional applications. Recently, the interest in sophisticated image analysis and recognition techniques as a way to enhance the built-in intelligence of systems has been greatly reduced in favour of new models of human perception, and advanced human-computer interaction tools aimed at exploiting the user's intelligence and understanding of the retrieval task at hand. A careful image domain and retrieval task analysis is also of great importance to ensure that queries are formulated at a semantic level appropriate for a specific application. A number of examples encompassing different semantic levels and application contexts, including retrieval of trademarks and of art images, are presented and discussed, providing insight into the state-of-the-art of content-based image retrieval systems and techniques.

1.2 IMAGE RETRIEVAL AND ITS APPLICATIONS

This Section includes a critical discussion of the main limitations affecting current CBIR systems, followed by a taxonomy of Visible Image Retrieval systems and applications from the perspective of semantic requirements.

1.2.1 Current Limitations of Content-Based Image Retrieval

Semantic gap. Due to the huge amount of heterogeneous information in modern digital archives, a common requirement for modern CBIR systems is that visual content annotation be automatic. This gives rise to a *semantic gap* – namely, a discrepancy between the query a user ideally *would* and the one which he actually *can* submit to an information retrieval system – limiting the effectiveness of image retrieval systems.

As an example of semantic gap in text-based retrieval, consider the task of extracting humorous sentences from a digital archive including books by Mark Twain: this is simply impossible to ask from a standard textual, syntactic database system. However, the same system will accept queries like “find me all the sentences including the word ‘steamboat’ ” without problems. Consider now submitting this last query (maybe using an example picture) to a current state-of-the-art, automatically-annotated image retrieval system including pictures from illustrated books of the 19th century: the system response will be unlikely to consist of a set of steamboat images. Current automatic annotations of visual content are in fact based on raw image properties, and all retrieved images will look like the example image with respect to their color, texture, etc. We can therefore conclude that the semantic gap is wider for images than for text: this is because, unlike text, images cannot be regarded as a syntactically structured collection of words, each with a well defined semantics. The word ‘steamboat’ stands for a thousand possible images of steamboats but, unfortunately, current visual recognition technology is very far from providing textual annotation – for example steamboat, river, crowd, etc. – of pictorial content.

First-generation CBIR systems were based on manual textual annotation to represent image content, thus exhibiting less evident semantic gaps than modern, automatic CBIR approaches. Manual/textual annotation proved to work reasonably well, for example, for newspaper photographic archives. However, this technique can only be applied to small data volumes and, to be truly effective, annotation must be limited to very narrow visual domains (e.g., photographs of buildings or of celebrities, etc.). Moreover, in some cases, textually annotating visual content can be a hard job (think, for example, of non-figurative graphic objects, such as trademarks). Note that the reverse of the sentence above seems equally true, namely, the image of a steamboat stands for a thousand words. Increasing the semantic level by manual intervention is also known to introduce subjectivity in the content classification process (going back to Mark Twain's example, one would hardly agree with the choice of humorous sentences made by the annotator). This can be a serious limitation, due to the difficulty of anticipating the queries that future users will actually submit.

The discussion above both provides insight into the semantic gap problem and suggests ways to solve it. Explicitly, *(i)* the notion of "information content" is extremely vague and ambiguous, as it reflects a subjective interpretation of data: there is no such thing as an objective annotation of information content, especially at a semantic level; *(ii)* nevertheless, modern CBIR systems are required to operate in an automatic way, and at a semantic level as close as possible to the one users are expected to refer to in their queries; *(iii)* gaps between system and user semantics are partially due to the nature of the information being searched, and partially due to the way a CBIR system operates; *(iv)* to bridge the semantic gap, extreme care should be devoted to the way CBIR systems internally represent visual information and externally interact with the users.

Recognition vs similarity retrieval. In the last few years, a number of CBIR systems using image recognition technologies proved reliable enough for professional applications in industrial automation, biomedicine, social security, etc. Face recognition systems are now widely used for biometric authentication and crime prevention [4]; similarly, automatic image-based detection of tumor cells in tissues is being used to support medical diagnosis and prevention [28].

However, there is much more to image retrieval than simple recognition. In particular, the fundamental role that human factors play in all phases of a CBIR project – from development to use – has been largely neglected in the CBIR literature. In fact, CBIR has long been considered only a sub-branch of consolidated disciplines such as pattern recognition, computer vision and even artificial intelligence, where interaction with a user plays a secondary role. To overcome some of the current limitations of CBIR, metrics, performance measures and retrieval strategies are now being developed which incorporate an active human participant in the retrieval process. Another distinction between recognition and retrieval is evident in less specialized domains, such as web search. These applications, among the most challenging for CBIR, are inherently concerned with ranking (i.e., re-ordering database images according to their measured similarity to a query example even if there is no image similar to the example) rather than classification (i.e., a binary partitioning process deciding whether or not an observed object matches a model), as the result of similarity-based retrieval.

	recognition	**similarity retrieval**
target performance	high precision	high recall, any precision
system output	database partition	database reordering/ranking
interactivity	low	high
user modeling	not important	important
built-in intelligence	high	low
application domain	narrow	wide
semantic level	high	application-dependent
annotation	manual	automatic
semantic range	narrow	wide
view invariance	yes	application-dependent

Table 1.1 Typical features of recognition and similarity retrieval systems (see text).

Image retrieval by similarity is the true distinguishing feature of a CBIR system, of which recognition-based systems should be regarded as a special case (see Table 1.1). Specifically, *(i)* The true qualifying feature of CBIR systems is the way human cooperation is exploited in performing the retrieval task; *(ii)* from the viewpoint of expected performance, CBIR systems typically require that all relevant images be retrieved, regardless of the presence of false positives (high recall, any precision); conversely, the main scope of image recognition systems is to exclude false positives, namely, to attain a high precision in the classification; *(iii)* recognition systems are typically required to be invariant w.r.t. a number of image appearance transformations (e.g., scale, illumination, etc.). In CBIR systems, it is normally up to the user to decide whether two images that differ, (say, with respect to color), should or should not be considered identical for the retrieval task at hand; *(iv)* as opposed to recognition, where uncertainties and imprecision are commonly managed automatically during the process, in similarity retrieval, it is the user who – being in the retrieval loop – analyzes system responses, refines the query, and determines relevance. This implies that the need for intelligence and reasoning capabilities inside the system is reduced. Image recognition capabilities, allowing the retrieval of objects in images much in the same way as words are found in a dictionary, are highly appealing to capture high level semantics and be used for the purpose of visual retrieval. However, it is evident from our discussion that CBIR typically requires versatility and adaptation to the user, rather than the embedded intelligence desirable in recognition tasks. Therefore, design efforts in CBIR are currently being devoted to combine lightweight, low semantics image representations with human-adaptive paradigms and powerful system-user interaction strategies.

1.2.2 Visible Image Retrieval Applications

Visible Image Retrieval (VisIR) can be defined as that branch of CBIR dealing with images produced with visible spectrum technology.

Since visible images are obtained through a large variety of mechanisms, including photographic devices, video cameras, imaging scanners, computer graphics software, etc., they are not expected to adhere to any particular technical standard of quality/resolution nor to any strict content characterization. In this chapter we focus on general-purpose systems for retrieval of photographic imagery.

Every CBIR application is characterized by a typical set of possible queries reflecting a specific semantic content. This Section classifies several important VisIR applications based on their semantic requirements; these are partitioned into three main levels.

Low-level. Here, the user's interest is concentrated on the basic perceptual features of visual content (dominant colors, color distributions, texture patterns, relevant edges and 2D shapes, uniform image regions) and on their spatial arrangement. Nearly all CBIR systems should support this kind of queries (cf. [8, 3]). Typical application domains for low-level queries are retrieval of trademarks and fashion design. Trademark image retrieval is useful to designers for the purpose of visual brainstorming, or to governmental organizations that need to check if a similar trademark already exists. Given the enormous number of registered trademarks (on the order of millions), this application must be designed to work fully automatically (actually, to date, in many European patent organizations trademark similarity search is still carried out in a manual way, through visual browsing). Trademark images are typically in black and white, but can also feature a limited number of unmixed and saturated colors and may contain portions of text (usually recorded separately). Trademarks symbols usually have a graphical nature, are only seldom figurative and often feature an ambiguous foreground/background separation. This is why it is preferable to characterize trademarks using descriptors such as color statistics and edge orientation [32, 12, 20].

Another application characterized by a low semantic level is fashion design: to develop new ideas, designers may want to inspect patterns from a large collection of images which look similar to a reference color and/or texture pattern. Low-level queries can support the retrieval of art images as well. For example, a user may want to retrieve all paintings sharing a common set of dominant colors or color arrangements, to look for commonalities and/or influences between artists with respect to the use of colors, spatial arrangement of forms and representation of subjects, etc. Of course, art images – as well as many other real application domains – encompass a range of semantic levels which go well beyond those provided by low-levels queries alone.

Intermediate-level. This level is characterized by a deeper involvement of users with the visual content. This involvement is peculiarly emotional, and is difficult to express in rational, textual terms. Examples of visual content with a strong emotional component can be derived from the visual arts (painting, photography). From the viewpoint of intermediate-level content, visual arts domains are characterized by the presence of either figurative elements like people, man-made objects, etc. or harmonic/disharmonic color contrast. Specifically, the shape of single objects dominates over color both in artistic photography (where, much more than color, concepts are conveyed through unusual views and details, and special effects such as motion blur) and in figurative art (of which R. Magritte is a noticeable example, since he combines painting techniques with photographic aesthetic criteria). Colors and color contrast between different image regions dominate shape in both mediaeval art and in abstract modern art (in both cases, emotions and symbols are predominant over verisimilitude). Art historians may be interested in finding images based on intermediate-level semantics. For example, they can consider the meaningful sensations that a painting provokes, according to the theory that different arrangements of colors on a canvas produces different psychological effects in the observer.

High-level. These are the queries which reflect data classification according to some rational criterion. For instance, journalism or historical image databases could be organized so as to be interrogated by genre (e.g., images of prime ministers, photos of environmental pollution, etc.). Other relevant applications fields range from advertising to home entertainment (e.g., management of family photo albums). Another example is encoding high-level semantics in the representation of art images, to be used by art historians, for example, for the purpose of studying visual iconography (see Sect. 1.4). State-of-the-art systems incorporating high-level semantics still require a huge amount of manual (and specifically textual) annotation, typically increasing with database size or task difficulty.

Web-search. Searching the web for images is one of the most difficult CBIR tasks. The web is not a structured database - its content is widely heterogeneous, and changes continuously.

Research in this area, although still in its infancy, is growing rapidly, with the goals of achieving high quality of service and effective search. An interesting methodology for exploiting automatic color-based retrieval to prevent access to pornographic images is reported in [15]. Preliminary image search experiments with a non-commercial system were reported in [30]. Two commercial systems, offering a limited number of search facilities, were launched in the past few years [13, 2]. Open research topics include: use of hierarchical organization of concepts and categories associated to visual content; use of simple but highly discriminant visual features, like color, so as to reduce the computational requirements of indexing; use of summary information for browsing and querying; use of analysis/retrieval methods in the compressed domain; and the use of visualization at different levels of resolution.

name	**low-level queries**	**advanced features**	**ref.**
Chabot	C	semantic queries	[24]
IRIS	C,T,S	semantic queries	[1]
MARS	C,T	user modeling, interactivity	[18]
NeTra	C,R,T,S	indexing, large databases	[21]
Photobook	S,T	user modeling, learning, interactivity	[25]
PICASSO	C,R,S	semantic queries, visualization	[8]
PicToSeek	C,R	invariance, WWW connectivity	[16]
QBIC	C,R,T,S,SR	indexing, semantic queries	[14]
QuickLook	C,R,T,S	semantic queries, interactivity	[5]
Surfimage	C,R,T	user modeling, interactivity	[23]
Virage	C,T,SR	semantic queries	[2]
Visual Retrievalware	C,T	semantic queries, WWW connectivity	[13]
VisualSEEk	R,S,SR	semantic query, interactivity	[29]
WebSEEk	C,R	interactivity, WWW connectivity	[30]

Table 1.2 Current retrieval systems (C=global color, R=color region, T=texture, S=shape, SR=spatial relationships). "Semantic queries" stands for queries at either intermediate- or high-level semantics (see text).

Despite the current limitations of CBIR technologies, several VisIR systems are available either as commercial packages or as free software on the web. Most of these systems are general purpose, even if they can be tailored to a specific application or thematic image collection, such as technical drawings, art images, etc. Some of the best-known VisIR systems are included in Tab. 1.2. The table reports both standard and advanced features for each system. Advanced features (to be further discussed in the following Sections) are aimed at complementing standard facilities, in order to provide enhanced data representations, interaction with users, or domain-specific extensions.

Unfortunately, most of the techniques implemented to date are still in their infancy.

1.3 ADVANCED DESIGN ISSUES

This Section addresses some advanced issues in visible image retrieval. As mentioned above, VisIR requires a new processing model in which incompletely specified queries are interactively refined, incorporating the user's knowledge and feedback to obtain a satisfactory set of results. Since the user is in the processing loop, the true challenge is to develop support for effective human-computer dialogue. This shifts the problem from putting intelligence in the system, as in traditional recognition systems, to interface design, effective indexing, and modeling of users' similarity perception and cognition. Indexing on the WWW poses additional problems concerned with the development of metadata for efficient retrieval and filtering.

Similarity modeling. Similarity modeling, a.k.a. user modeling, requires internal image representations that closely reflect the ways in which users interpret, understand and encode visual data. Finding suitable image representations based on low-level, perceptual features – like color, texture, shape, image structure, spatial relationships – is an important step toward the development of effective similarity models, and has been an intensively studied CBIR research topic in the last few years. Yet, using image analysis and pattern recognition algorithms to extract numeric descriptors which give a quantitative measure of perceptual features is only part of the job: many of the difficulties still remain to be addressed. In several retrieval contexts, higher level semantic primitives such as objects or even emotions induced by visual material should also be extracted from images and represented in the retrieval system, since it is these higher level features which – as semioticians and psychologists suggest – actually convey meaning to the observer (colors, for example, may induce particular sensations according to their chromatic properties and spatial arrangement). Of course, when direct manual annotation of image content is not possible, embedding higher level semantics into the retrieval system must follow from reasoning about perceptual features themselves.

A process of semantic construction driven by low level features and suitable for both advertising and artistic visual domains was recently proposed in [7] (see also Sect. 1.4). The approach characterizes visual meaning through an hierarchy, where each level is connected to its ancestor by a set of rules obtained through a semiotic analysis of the visual domains studied.

It is important to note that completely different representations can be built starting from the same basic perceptual features: it all depends on the intepretation of the features themselves. For instance, color-based representations can be more or less effective in terms of human similarity judgement depending on the color space used.

Also of crucial importance in user modeling is the design of similarity metrics used to compare current query and database feature vectors. In fact, human similarity perception is based on the measurement of an appropriate distance in a *metric psychological space*, whose form is doubtless quite different from the metric spaces (such as the Euclidean) typically used for vector comparison. Hence, to be truly effective, feature representation and feature matching models should somehow replicate the way in which humans assess similarity between different objects. This approach is complicated by the fact that there is no single model of human similarity. In [27], various definitions of similarity measures for feature spaces are presented and analyzed, with the purpose of finding characteristics of the distance measure which are relatively independent of the choice of the feature space.

System adaptation to individual users is another hot research topic. In the traditional approach of querying by visual example, the user explicitly indicates which features are important, selects a representation model, and specifies the range of model parameters and the appropriate similarity measure. Some researchers have pointed out that this approach is not suitable for general databases of arbitrary content or for average users [25]. It is instead suitable to domain-specific retrieval applications, where images belong to an homogeneous set, and users are experts. In fact, it requires that the user be aware of the effects of the representation and similarity processing on retrieval. A further drawback to this approach is its failure to model user's subjectivity in similarity evaluation. Combining multiple representation models can partially resolve this problem. If the retrieval system allows multiple similarity functions, the user should be able to select those that most closely model his/her perception.

Learning is another important way to address similarity and subjectivity modeling. The system presented in [22] is probably the best-known example of subjectivity modeling through learning. Users can define their subjective similarity measure through selections of examples and by interactively grouping similar examples. Similarity measures are obtained not by computing metric distances, but as a compound grouping of pre-computed hierarchy nodes. The system also allows manual and automatic image annotation through learning, by allowing the user to attach labels to image regions. This permits semantic groupings and the usage of textual keys for querying and retrieving database images.

Interactivity. Interfaces for content-based interactivity provide access to visual data by allowing the user to switch back and forth between navigation, browsing and querying. While querying is used to precisely locate certain information, navigation and browsing support exploration of visual information spaces. Flexible interfaces for querying and data visualization are needed to improve the overall performance of a CBIR system. Any improvement in

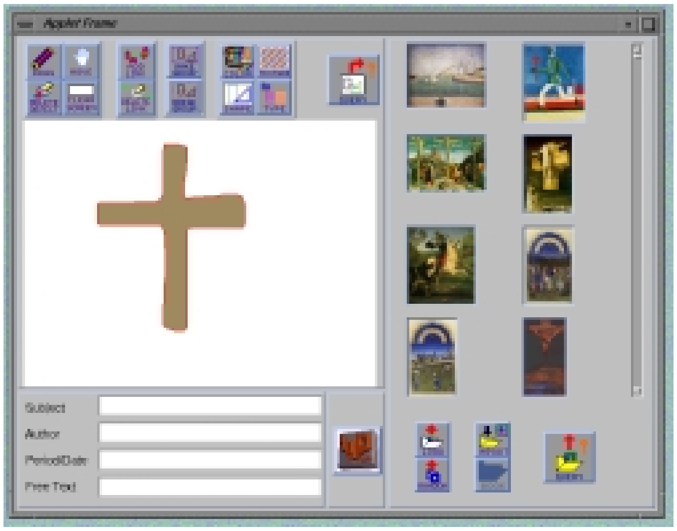


Fig. 1.1 Image retrieval with conventional interaction tools: query space and retrieval results (thumbnail form).

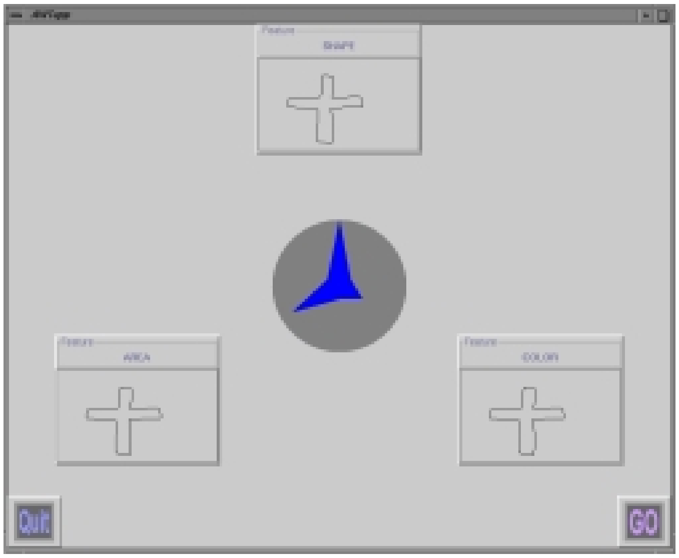


Fig. 1.2 Image retrieval with advanced interaction tools: query composition in "star" form (see text).

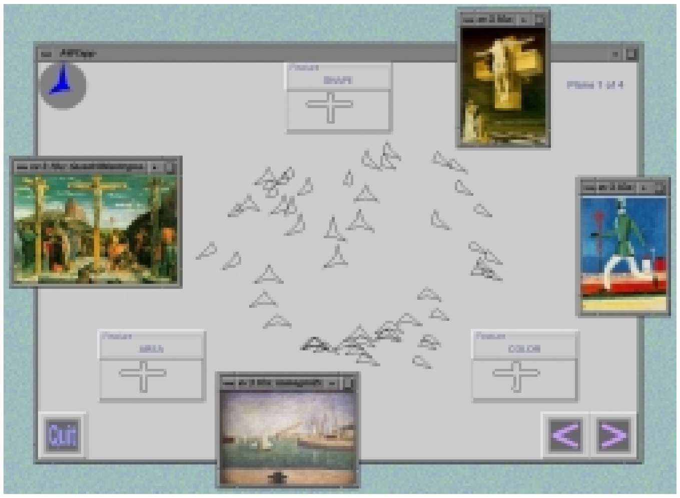


Fig. 1.3 Image retrieval with advanced interaction tools: result visualization in "star" form (see text).

interactivity, while pushing towards a more efficient exploitation of human resources during the retrieval process, also proves particularly appealing for commercial applications supporting non-expert (hence more impatient and less adaptive) users. Often a good interface can let the user express queries which go beyond the normal system representation power, giving the user the impression of working at a higher semantic level than the actual one. As an example, sky images can be effectively retrieved by a blue color sketch in the top part of the canvas; similarly, "all leopards" in an image collection can be retrieved by querying for texture (possibly invariant to scale), using a leopard's coat as an example.

There is a need for query technology that will support more effective ways to express composite queries, thus combining high-level textual queries with queries by visual example (icon, sketch, painting, whole image). In retrieving visual information, high-level concepts, such as the type of an object, or its role if available, are often used together with perceptual features in a query; yet, most current retrieval systems require the use of separate interfaces for text and visual information. Research in data visualization can be exploited to define new ways of representing the content of visual archives and the paths followed during a retrieval session. For example, new effective visualization tools have recently been proposed which enable the display of whole visual information spaces instead of simply displaying a limited number of images [17].

Fig. 1.1 shows the main interface window of a prototype system allowing querying by multiple features [9]. In the figure, retrieval by shape, area, and color similarity of a cross-like sketch is supported with a very intuitive mechanism, based on the concept of "star." Explicitly, an n-point star is used to perform an n-feature query, the length of each star point being proportional to the relative relevance of the feature with which it is associated. The relative weights of the three query features is indicated by the 3-point star shown at query composition time (Fig. 1.2): an equal importance is assigned to shape and area, while a lesser importance is assigned to color. Displaying the most relevant images in thumbnail format is the most common method to present retrieval results (see again Fig. 1.1). Display of thumbnails is usually accompanied by display of the query, so that the user can visually compare retrieval results with the original request, and provide relevance feedback accordingly [26]. However, thumbnail display has several drawbacks: *(i)* thumbnails must be displayed on a number of successive pages (each page containing a maximum of, say, twenty thumbnails); *(ii)* for multiple-feature queries, the criteria for ranking the thumbnail images is not obvious; *(iii)* comparative evaluation of relevance is difficult, and usually limited to thumbnails in the first one or two pages.

A more effective visualization of retrieval results is therefore suggested. Fig. 1.3 shows a new visualization space which displays retrieval results in star form rather than in thumbnail form. This representation is very useful for compactly describing the individual similarity of each each image w.r.t. the query, as well as how images sharing similar features are distributed inside the database. In the example provided, which refers to the query of Figs. 1.1–1.2, stars located closer to the center of the visualization space have a higher similarity w.r.t. the query (the first four of them are reported at the sides of the visualization space). Images at the bottom center of the visualization space are characterized by a good similarity w.r.t. the query in terms of area and color, but their shape is quite different from that of the query. This method of visualizing results permits an enhanced user-system synergy for the progressive refinement of queries, and allows for a degree of uncertainty in both the user's request and the content description. In fact, the user is able to refine his query by a simple change in the shape of the query star, based on the shape of the most relevant results obtained in the previous iteration.

Another useful method for narrowing the semantic gap between system and user is to provide the user with a visual interpretation of the internal image representation that allows them to refine/modify the query [6]. Fig. 1.4 shows how the original external query image is transformed into its internal counterpart through a multiple-region content representation based on color histograms. The user is able to refine the original query by directly reshaping the single histograms extracted from each region and examining how this affects the visual appearance of the internal query; the latter – and not the external query – is the one actually used for similarity matching inside the database.

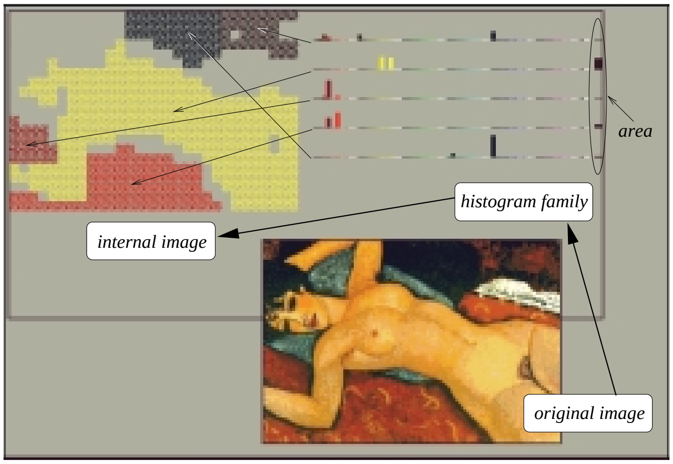
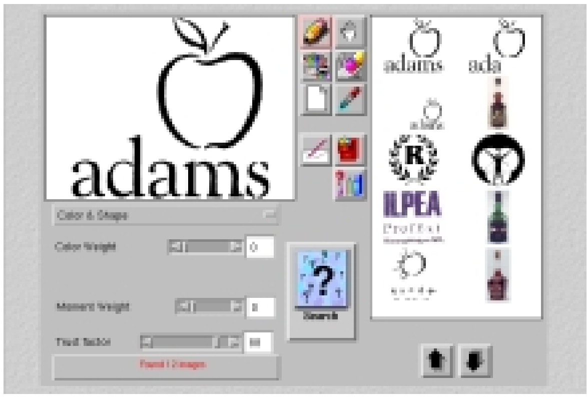


Fig. 1.4 Visualization of internal query representation.


Color & Shape

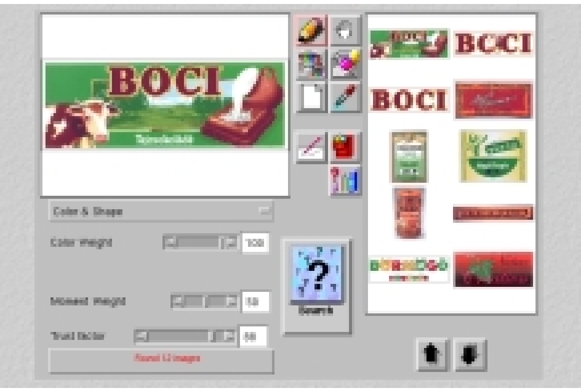
Color Weight 0

Moment Weight 8

Trust factor 88

Found 12 images

Fig. 1.5 Retrieval of trademarks by shape only.



Color & Shape

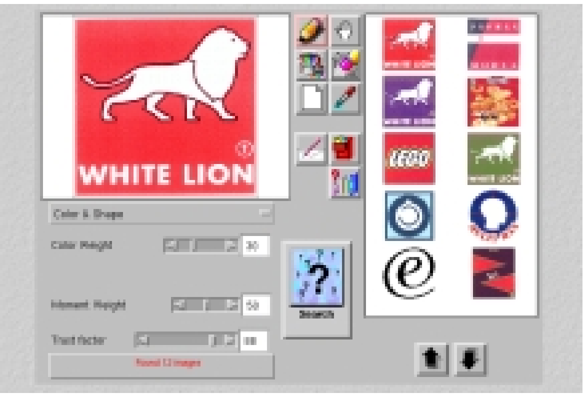
Color Weight 100

Moment Weight 50

Trust factor 88

Found 12 images

Fig. 1.6 Retrieval of trademarks by color only.



Color & Shape

Color Weight 30

Moment Weight 50

Trust factor 88

Found 12 images

Fig. 1.7 Retrieval of trademarks by combined shape and color.

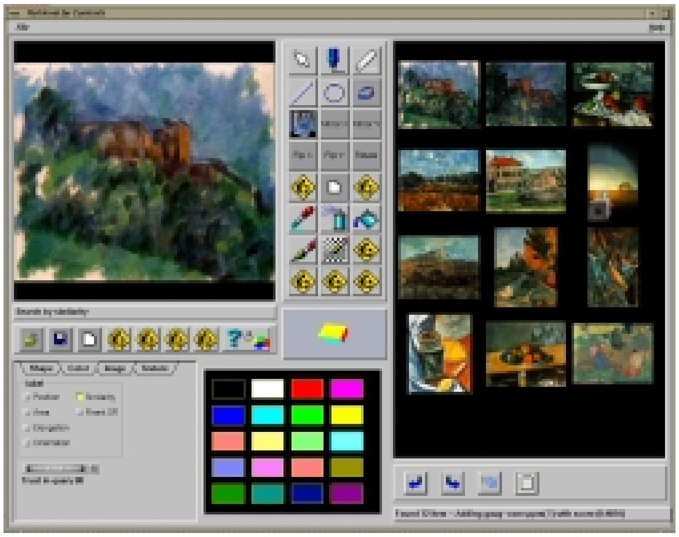


Fig. 1.8 Retrieval of art paintings by color similarity.

1.4 VISIBLE IMAGE RETRIEVAL EXAMPLES

This Section shows several examples of image retrieval using packages developed at the Visual Information Laboratory of the University of Florence [8]. These packages include a number of advanced retrieval features. Some of these features have been outlined above, and are also present in several other available VisIR packages. The examples are provided in increasing order of representation complexity (semantic demand), ranging from trademark through art images and iconography. For the purpose of efficient system design, image representation was designed to support the most common query types, which in general are strictly related to how images are used in the targeted application domain.

Retrieval of trademarks by low-level content. Due to domain characteristics of trademark images, the representation used in this case is based on very simple perceptual features, namely, edge orientation histograms and their moments to represent shape, and color histograms. Image search can be based on color and shape taken in any proportion. Shape moments can be excluded from the representation to enable invariance w.r.t. image size.

Figs. 1.5 through 1.7 show three different retrieval tasks from an experimental database of a thousand entries (in all cases, the example is in the upper left window of the interface). Fig. 1.5 shows the result in the case of retrieval by shape only: notice that, beside being totally invariant to scale (see the 3rd ranked image), the chosen representation is also partially invariant to partial

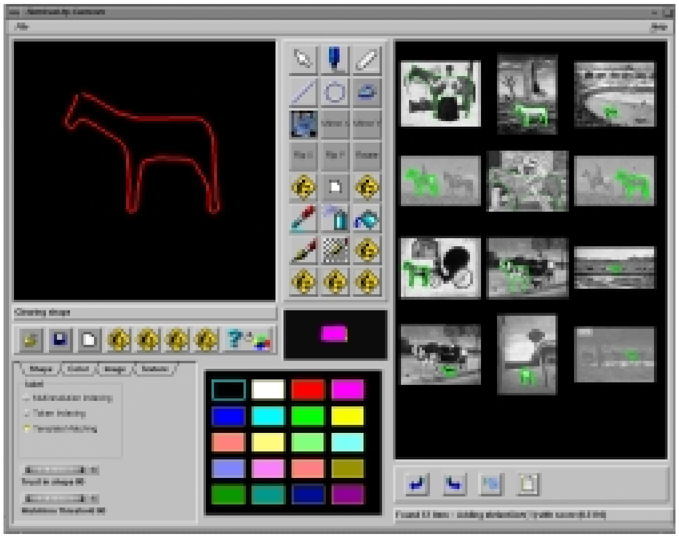


Fig. 1.9 Retrieval of art paintings by shape similarity.

writing changes. Retrieval by color only is shown in Fig. 1.6; the trademarks retrieved all contain at least one of the two dominant colors of the example. The third task, shown in Fig. 1.7, is to perform retrieval based on both color and shape, shape being dominant to color. All trademarks with the white lion were correctly retrieved, regardless of the background color.

Retrieval of paintings by low- and intermediate-level content. The second example demonstrates retrieval from an experimental database featuring hundreds of modern art paintings. Both low- and intermediate-level queries are supported. From our discussion, it is apparent that color and shape are the most important image characteristics for feature-based retrieval of paintings. Image regions are extracted automatically by means of a multiresolution color segmentation technique, based on an energy-minimization process. Chromatic qualities are represented in the $L^*u^*v^*$ space, to gain a good approximation of human color perception, and similarity of color regions is evaluated considering both chromatic and spatial attributes (region area, location, elongation, orientation) [10]. A more sophisticated color representation is required than with trademarks, due to the much more complex color content of art images. The multiresolution strategy adopted allows the system to take into account color regions scattered throughout an image. Fig. 1.8 shows color similarity retrieval results using a painting by P. Cezanne as the query image. Notice how many of the retrieved images are actually paintings by the same painter; this is sensible, as it reflects the preference of each artist for specific color combinations.

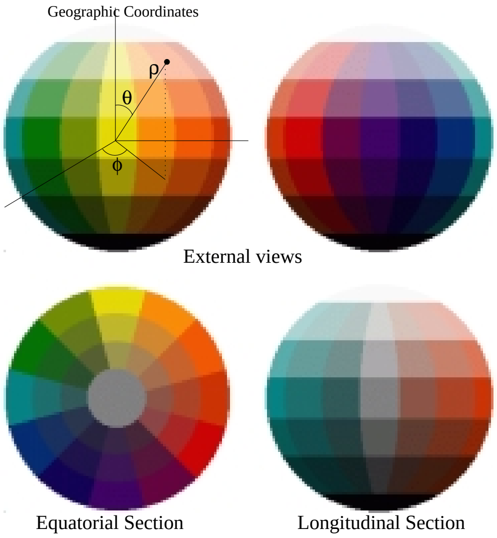


Fig. 1.10 The Itten sphere.

Due to their too high semantic level, objects are annotated manually (but not textually) in each image by object contour drawing. For the purpose of shape-based retrieval, queries are submitted by sketch; query and database shapes are compared through an energy-minimization procedure, where the sketch is elastically deformed to best fit the target shape [11]. Querying by sketch typically gives the user the (false, but pleasant) impression that the system is more "intelligent" than it really is; in fact, the system would be unable to extract an object shape from an example query image without the manual drawing made by the user. Results of retrieval by shape are shown in Fig. 1.9, in response to a horse query. Many of the retrieved images actually include horses or horse-like figures.

Retrieval of paintings by semiotic content. As a second example of the way intermediate-level content can be represented and used, this Section reviews the approach recently proposed in [7] for enhancing the semantic representation level for art images according to semiotic principles. In this approach, a content representation is built through a process of syntactic construction, called "compositional semantics," featuring the composition of higher semantic levels according to syntactic rules operating at a perceptual feature level. The rules are directly translated from the aesthetic and psychological theory of J. Itten [19] on the use of color in art and the semantics that it induces. Itten observed that color combinations induce effects such as harmony, disharmony, calmness, excitement, which are consciously exploited by artists in the composition of their paintings. Most of these effects are related to high-level

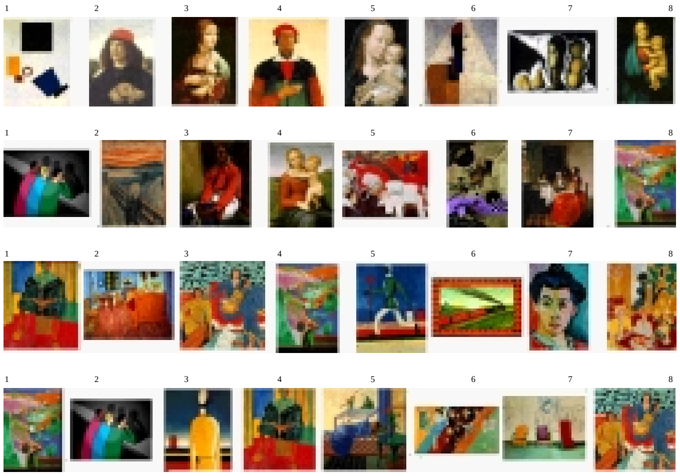


Fig. 1.11 Best ranked images according to queries for contrast of luminance (top row), contrast of saturation (second row), contrast of warmth (third row) and harmonic accordance (bottom row).

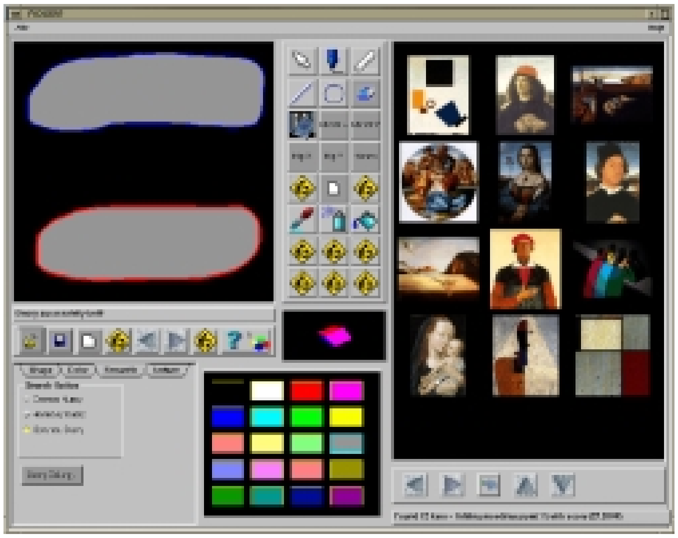


Fig. 1.12 Results of a query for images with two large regions with contrasting luminance

chromatic patterns rather than to physical properties of single points of color. The theory characterizes colors according to the categories of *hue*, *luminance* and *saturation.* Twelve hues are identified as fundamental colors, and each fundamental color is varied through five levels of luminance and three levels of saturation. These colors are arranged into a chromatic sphere, such that perceptually contrasting colors have opposite coordinates w.r.t. the center of the sphere (Fig. 1.10). Analyzing the polar reference system, four different types of contrasts can be identified: contrast of *pure colors, light–dark, warm–cold, quality (saturated–unsaturated).* Psychological studies have suggested that, in western culture, red–orange environments induce a sense of warmth (yellow through red–purple are *warm* colors). Conversely, green-blue conveys a sensation of cold (yellow–green through purple are *cold* colors). Cold sensations can be emphasized by the contrast with a warm color or damped by its coupling with a highly cold tint. The term *harmonic accordance* refers to combinations of hue and tone that are pleasing to the human eye. Harmony is achieved by the creation of color combinations, selected by connecting locations through regular polygons inscribed within the chromatic sphere.

Fig. 1.11 shows the eight best ranked images retrieved by the system in response to four reference queries, addressing contrasts of luminance, warmth, saturation and harmony, respectively. Tests show good agreement between human opinions (from interviews) and the system in the assignment of similarity rankings [7]. Fig. 1.12 shows an example of retrieval of images characterized by two large regions with contrasting luminance, from a database of several hundred XV- to XX-century paintings. Two dialog boxes are used to define properties (hue and dimension) of the two sketched regions of Fig. 1.12. Retrieved paintings are shown in the right part of Fig. 1.12. The twelve best-matched images all display a relevant luminance contrast, featuring a black region over a white background. Images ranked in the second, third, and fifth through seventh positions are all examples of how contrast of luminance between large regions can be used to convey the perception of different planes of depth.

Retrieval of Renaissance paintings by low- and high-level content. Iconographic study of Renaissance paintings provides an interesting example of simultaneous exploitation of low- and high-level descriptors [31]. In this retrieval example, spatial relationships and other features like color/texture are combined with textual annotations of visual entities. Modeling of spatial relationships is obtained through an original modeling technique which is able to account for the overall distribution of relationships among the individual pixels belonging to the two regions. Textual labels are associated with each manually marked object (in the case of Fig. 1.13, these are “Madonna” and “angel”). The spatial relationship between an observing and an observed object is represented by a finite set of equivalence classes (the *symbolic walk-throughs*) on the sets of possible paths leading from any pixel in the observing object to any pixel in the observed object. Each equivalence class is asso-

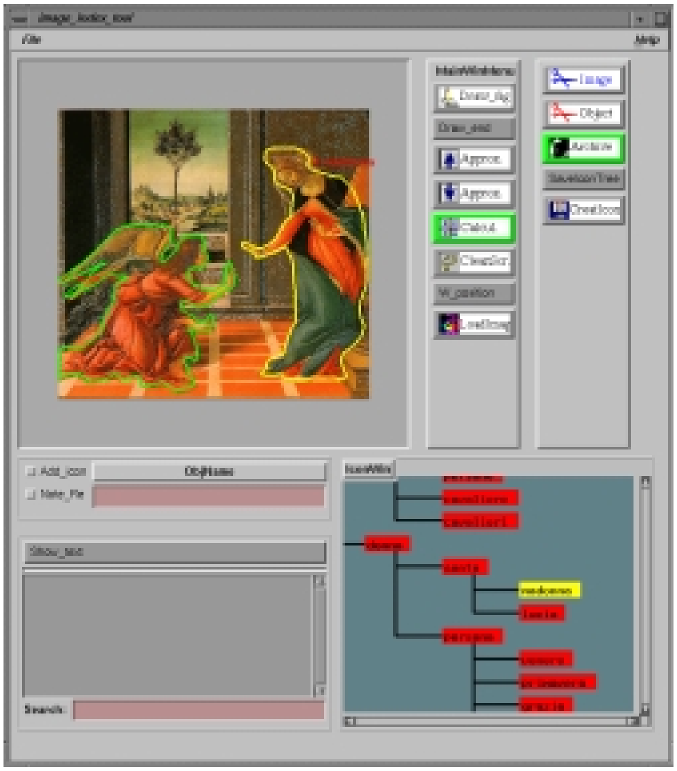


Fig. 1.13 Manual annotation of image content through graphics and text.

ciated with a weight which provides an integral measure of the set of pixel pairs that are connected by a path belonging to the class, thus accounting for the degree to which the individual class represents the actual relationship between the two regions. The resulting representation is referred to as a *weighted walkthroughs* model. Art historians can, for example, perform iconographic search by finding, for example, all paintings featuring the Madonna and another figure in a desired spatial arrangement. (in the query of Fig. 1.14 *left*, the configuration is that of a famous annunciation). Retrieval results are shown in Fig. 1.14: Note that the top-ranked images all depict annunciation scenes where the Madonna is on the right side of the image. Due to the strong similarity in the spatial arrangement of figures – spatial arrangement has a more relevant weight than figure identity in this example – non-annunciation paintings including the Madonna and a saint are also retrieved.

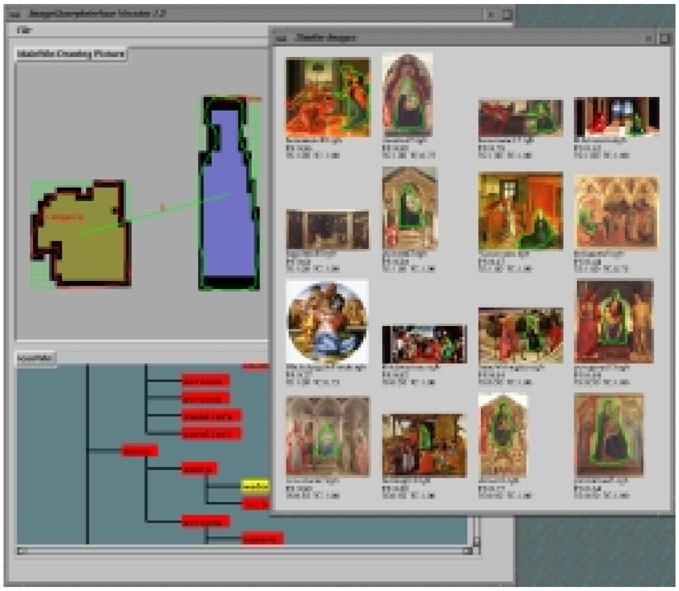


Fig. 1.14 Iconographic search: query submission and retrieval results.

1.5 CONCLUSION

In this chapter, a discussion about current and future issues in content-based visible image retrieval design was presented with an eye to applications. The ultimate goal of a new-generation visual retrieval system is to achieve fully automatic annotation of content and reduce the semantic gap by skillfully exploiting user's intelligence and objectives. This can be obtained by stressing the aspects of human similarity modeling and user-system interactivity. The discussion and the concrete retrieval examples illustrate that, despite the huge efforts made in the last few years, research on visual retrieval is still in its infancy. This is particularly true for applications intended not for professional/specialist use, but for the mass market, namely, for naive users. Designing effective retrieval systems for general use is a big challenge that will no doubt require extra research efforts to make systems friendly and usable, but will also open new markets and perspectives in the field.

Acknowledgements

The authors would like to thank the editors, L. Bergman and V. Castelli, for their kindness and support. G. Baldi, S. Berretti, P. Pala and E. Vicario from the Visual Information Laboratory of the University of Florence provided the example images shown in Section 1.4: thanks to all.

References

1. P. Alshuth, T. Hermes, C. Klauck, J. Kreiss and M. Roper, “IRIS Image Retrieval for Images and Video,” Proc First Int’l Workshop on Image Database and Multi-media Search, 1996.

2. J.R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horovitz, R. Humphrey and R. Jain, “The Virage Image Search Engine: an Open Framework for Image Management,” in *Proc. SPIE Int’l Conf. on Storage and Retrieval for Still Image and Video Databases*, 1996.

3. E. Binaghi, I. Gagliardi and R. Schettini, “Image Retrieval Using Fuzzy Evaluation of Color Similarity,” *Int. Journal of Pattern Recognition and Artificial Intelligence* 8(4):945–968,1994.

4. R. Chellappa, C.L. Wilson and S. Sirohey, “Human and Machine Recognition of Faces: A Survey,” *Proceedings of the IEEE* 83(5):705–740, 1995.

5. G. Ciocca, R. Schettini, “A Relevance Feedback Mechanism for Content-Based Image Retrieval,” *Information Processing and Management*, 35:605–632, 1999.

6. C. Colombo and A. Del Bimbo, “Color-Induced Image Representation and Retrieval,” *Pattern Recognition* 32:1685–1695, 1999.

7. C. Colombo, A. Del Bimbo, and P. Pala, “Semantics in Visual Information Retrieval,” *IEEE Multimedia* 6(3):38–53, 1999.

8. A. Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann, San Francisco, CA, 1999.

9. A. Del Bimbo and P. Pala, “Image Retrieval by Multiple Features Combination,” *Technical Note*, Dept. of Systems and Informatics, University of Florence, Italy, 1999.

10. A. Del Bimbo, M. Mugnaini, P. Pala and F. Turco, “Visual Querying by Color Perceptive Regions,” *Pattern Recognition* 31(9):1241–1253, 1998.

11. A. Del Bimbo and P. Pala, “Retrieval by Elastic Matching of User Sketches,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19(2):121–132, 1997.

12. J.P. Eakins, J.M. Boardman, and M.E. Graham, “Similarity Retrieval of Trade Mark Images,” *IEEE Multimedia* 5(2):53–63, 1998.

13. J. Feder, “Towards Image Content-Based Retrieval for the World-Wide Web,” *Advanced Imaging* 11(1):26–29, 1996.

14. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Janker, “Query by Image and Video Content: the QBIC System,” *IEEE Computer* 28(9):310–315,1995.

15. D. Forsyth, M. Fleck and C. Bregler, “Finding Naked People,” in *Proc. European Conf. on Computer Vision*, 1996.

16. T. Gevers and A.W.M. Smeulders, “The PicToSeek WWW Image Search System,” in *Proc. Int'l Conf. on Multimedia Computing and Systems*, 1999.

17. A. Gupta, S. Santini and R. Jain, “In Search of Information in Visual Media,” *Comm. of the ACM* 40(12):35–42, 1997.

18. T. Huang et al., “Multimedia Analysis and Retrieval System (MARS) Project,” in *Digital Image Access and Retrieval*, P.B. Heidorn and B. Sandore eds., 1997.

19. J. Itten, *Kunst der Farbe*, Otto Maier Verlag, Ravensburg, Germany, 1961 (in German).

20. A.K. Jain and A. Vailaya, “Shape-based Retrieval: a Case Study with Trademark Image Database,” *Pattern Recognition* 31(9):1369–1390, 1998.

21. W.-Y. Ma and B.S. Manjunath, “NeTra: A Toolbox for Navigating Large Image Databases,” *Multimedia Systems* 7:184–198, 1999.

22. T. Minka and R. Picard, “Interactive Learning with a Society of Models,” *Pattern Recognition* 30(4):565–582, 1997.

23. C. Nastar et al., “Surfimage: A Flexible Content-Based Image Retrieval System,” in *Proc. ACM Multimedia*, 1998.

24. V.E. Ogle and M. Stonebraker, “Chabot: Retrieval from a Relational Database of Images,” *IEEE Computer* 28(9):40–48, 1995.

25. R. Picard, T.P. Minka and M. Szummer, “Modeling User Subjectivity in Image Libraries,” in *Proc. IEEE Int'l Conf. on Image Processing*, 1996.

26. Y. Rui, T.S. Huang, M. Ortega and S. Mehotra, “Relevance Feedback: A Powerful Tool for Interactive Content-Based Image Retrieval,” *IEEE Trans. on Circuits and Systems for Video Technology* 8(5):644–655, 1998.

27. S. Santini and R. Jain, "Similarity Measures," *IEEE Trans. on Pattern Analysis and Machine Intelligence* 21(9):871–883, 1999.

28. C.R. Shyu, C.E. Brodley, A.C. Kak, A. Kosaka, A.M. Aisen and L.S. Broderick, "ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases," *Computer Vision and Image Understanding* 75(1/2):175–195, 1999.

29. J.R. Smith and S.-F. Chang, "Querying by Color Regions Using the VisualSEEk Content-Based Visual Query System," in *Intelligent Multimedia Information Retrieval*, M.T. Maybury, ed., 1997.

30. S.-F. Chang, J.R. Smith, M. Beigi and A. Benitez, " Visual Information Retrieval from Large Distributed Online Repositories," *Comm. of the ACM* 40(12):63–71, 1997.

31. E. Vicario and He Wengxe, "Weighted Walkthroughs in Retrieval by Content of Pictorial Data," in *Proc. IAPR-IC Int'l Conf. on Image Analysis and Processing*, 1997.

32. J.K. Wu, C.P. Lam, B.M. Metre, Y.J. Gao and A.D. Narasimhalu, "Content-Based Retrieval for Trademark Registration," *Multimedia Tools and Applications* 3(3):245–267, 1996.