10

Color Spatial Arrangement for Image Retrieval by Visual Similarity

Stefano Berretti and Alberto Del Bimbo

CONTENTS

10.1 Introduction

The rapid advancements in multimedia technology have increased the relevance that repositories of digital images are assuming in a wide range of information systems. Effective access to such archives requires that conventional searching techniques based on external textual keywords be complemented by content-based queries addressing appearing visual features of searched data [1], [2]. To this end, a number of models were experimented with that permit the representation and comparison of images in terms of quantitative indexes of visual features [3], [4], [5]. In particular, different techniques were identified and experimented with to represent the content of single images according to low-level features, such as color [6], [7], [8], texture [9], [10], shape [11], [12], [13], and structure [14], [15];

intermediate-level features of saliency [16], [17], [18] and spatial relationships [19], [20], [21], [22], [23]; or high-level traits modeling the semantics of image content [24], [25], [26]. In doing so, extracted features may either refer to the overall image (e.g., a color histogram), or to any subset of pixels constituting a spatial entity with some apparent visual cohesion in the user's perception. This can be the set of pixels constituting any object with high-level semantics, such as a character, a face, or a geographic landmark. Or it can be a set of pixels with low-level visual cohesion, induced by a common chrominance or texture, or by a common position within a predefined area of the image. As a limit case, the overall image can be regarded as a particular spatial entity.

Selecting the entities on which content representation should be based entails a trade-off between the significance of the model and the complexity of its creation: models containing high-level entities permit a closer fit to the users' expressive habits, but they also require some assistance in the archiving stage for the identification and the classification of significant entities.

Information associated with each entity generally combines a set of salient entity features, along with additional indexes that can be measured once the entity has been extracted: a high-level object is usually associated with a symbolic type [19], [27], an image region derived through a color-based segmentation is associated with a chromatic descriptor [28], and both of them can be associated with a measure of size, or with any other shape index [29], [30], [31]. When multiple entities are identified, the model may also capture information about their mutual spatial relationships. This can improve the effectiveness of retrieval by registering perceived differences and similarities that depend on the arrangement of entities rather than on their individual features. Relational information associated with multiple entities can capture high-level concepts, such as an action involving represented objects or spatial relationships between the pixel sets representing different entities. Relationships of the latter kind are most commonly employed in content-based image retrieval (CBIR) due to the possibility of deriving them automatically and to their capability of conveying a significant semantics.

In particular, image representations based on chromatic indexes have been widely experimented and comprise the basic backbone of most commercial and research retrieval engines, such as QBIC [32], Virage [33], VisualSeek [20], PickToSeek [34], BlobWorld [35], and SIMPLIcity [36], [37], to mention a few. This apparently depends on the capability of color-based models to combine robustness of automatic construction with a relative perceptual significance of the models.

However, despite the increased descriptive capability enabled by relational models that identify separate spatial entities, in the early and basic approaches, the chromatic content of the overall image has been represented by a global histogram. This is obtained by tessellating the (three-dimensional) space of colors into a finite set of reference parts, each associated with a bin representing the quantity of pixels with color that belongs to the part itself [38]. The similarity between two images is thus evaluated by comparing bins and their distribution [39]. In doing so, the evaluation of similarity does not account for the spatial arrangement and coupling of colors over the image. This plays a twofold role in the user's perception, serving to distinguish images with common colors and to perceive similarities between images with different colors but similar arrangements. To account for both these aspects, chromatic information must be associated with individual spatial entities identified over the image. According to this, integration of spatial descriptors and color has been addressed to extend the significance of color histograms with some index of spatial locality.

In early work [40], the image is partitioned into blocks along a fixed grid, and each block is associated with an individual local histogram. In this case, similarity matching also considers adjacency conditions among blocks with similar histograms. However, because blocks are created according to a static partitioning of the image, representation of spatial arrangement does not reflect the user-perceived patching of colors. In Reference [41], the spatial arrangement of the components of a color histogram is represented through color correlograms, capturing the distribution of distances between pixels belonging to different bins. In Reference [28], a picture is segmented into color sets and partitioned into a finite number of equally spaced slices. The spatial relationship between two color sets is modeled by the number of slices in which one color set is above the other. In Reference [31], the spatial distribution of a set of pixel blocks with common chromatic features is indexed by the two largest angles obtained in a Delaunay triangulation over the set of block centroids. Though quantitative, these methods still do not consider the actual extensions of spatial entities.

To overcome the limit, the image can be partitioned into entities collecting pixels with homogeneous chromatic content [42]. This can be accomplished through an automated segmentation process [43], [44], which clusters color histograms around dominating components, and then determines entities as image segments collecting connected pixels under common dominating colors [45], [46], [47], [48]. In general, few colors are sufficient to partition the histogram in cohesive clusters, which can be represented as a single average color without significant loss for the evaluation of similarity. However, color clusters may be split into several nonconnected image segments when they are back-projected from the color space to the image. This produces an exceedingly complex model, which clashes with the human capability to merge regions with common chromatic attributes. An effective solution to this problem was proposed in References [21], [49], where weighted walkthroughs are proposed to quantitatively model spatial relationships between nonconnected clusters of color in the image plane. Following a different approach, in Reference [50], spatial color distribution is represented using local principal component analysis (PCA). The representation is based on image windows that are selected by a symmetry-based saliency map and an edge and corner detector. The eigenvectors obtained from local PCA of the selected windows form color patterns that capture both low and high spatial frequencies, so they are well suited for shape as well as texture representation.

To unify efforts aiming to define descriptors that effectively and efficiently capture the image content, the International Standards Organization (ISO) has developed the MPEG-7 standard, specifically designed for the description of multimedia content [51], [52], [53]. The standard focuses on the representation of descriptions and their encoding, so as to enable retrieval and browsing applications without specific ties to a single content provider. According to this, descriptors are standardized for different audiovisual features, such as dominant color, texture, object's contour shape, camera motion, and so forth. (All MPEG-7 descriptors are outlined in Reference [54].) This has permitted research efforts to focus mainly on optimization mechanisms rather than on the definition and extraction of the descriptors. In particular, CBIR applications have usefully exploited the features provided by the standard. For example, solutions like those proposed in References [55, 56] have tried to combine MPEG-7 descriptors with relevance feedback mechanisms [57] in order to improve the performances of retrieval systems. In other works, because the MPEG-7 does not standardize ways whereby content descriptions should be compared, effective models for evaluating similarities among descriptors have been investigated [58].

In these approaches, chromatic descriptors are widely used. Specifically, MPEG-7 provides seven color descriptors, namely, *Color space, Color Quantization, Dominant Colors, Scalable Color, Color Layout, Color-Structure*, and *Group of Frames/Group of Pictures Color*. Among these, the color layout descriptor (CLD) and the color-structure descriptor (CSD) are capable of conveying spatial information of the image color distribution. The CSD provides information regarding color distribution as well as localized spatial color structure in the image. This is obtained by taking into account all colors in a structuring element of 8 × 8 pixels that slides over the image, instead of considering each pixel separately. Unlike the color histogram, this descriptor can distinguish between two images in which a given color is present in identical amounts, but where the structure of the groups of pixels having that color is different. Information carried out by the CSD are complemented using the CLD, which provides information about the color spatial distributions by dividing images into 64 blocks and extracting a representative color from each of the blocks to generate an 8 × 8 icon image. When regions are concerned, the region locator descriptor (RLD) can be used to enable region localization within images by specifying them with a brief and scalable representation of a box or a polygon.

However, these kinds of descriptors permit some information to be embedded on the spatial localization of color content into color histograms but may not be appropriate for capturing binary spatial relationships between complex spatial entities. For example, this is the case in which users are interested in retrieving images where several entities, identified either by high-level types or low-level descriptors, are mutually positioned according to a given pattern of spatial arrangement. Moreover, this difficulty is particularly evident in expressing spatial relationships between nonconnected entities.

In this chapter, we propose an original representation of the spatial arrangement of chromatic content that contributes to the state-of-the-art in two main respects. First, the color information is captured by partitioning the image space in color clusters collecting pixels with common chromatic attributes, regardless of their spatial distribution in separate segments. This improves perceptual robustness and facilitates matching and indexing. In particular, this avoids the excessive complexity of descriptions arising in segmenting images based on connected regions of homogeneous color. However, it also poses some major difficulties related to the spatial complexity of the projection of color clusters and to the consequent difficulty in representing their arrangement. To this end, as a second contribution of this work, we propose and expound a descriptor, called *weighted walkthroughs*, that is able to capture the binary directional relationship between two complex sets of pixels, and we embed it into a graph theoretical model. In fact, weighted walkthroughs enable a quantitative representation of the joint distribution of masses in two extended spatial entities. This relationship is quantified over the dense set of pixels that comprise the two entities, without reducing them to a minimum embedding rectangle or to a finite set of representative points. This improves the capability to discriminate perceptually different relationships and makes the representation applicable for complex and irregular-shaped entities. Matching a natural trait of vagueness in spatial perception, the relationship between extended entities is represented as the union of the primitive directions (the walkthroughs) which connect their individual pixels. The mutual relevance of different directions is accounted for by quantitative values (the weights) that enable the establishment of a quantitative metric of similarity. Breaking the limits of Boolean classification of symbolic models, this prevents classification discontinuities and improves the capability to assimilate perceptually similar cases. Weights are computed through an integral form that satisfies a main property of compositionality. This permits efficient computation of the relationships between two entities by linear combination of the relationships between their parts, which is not possible for models based on symbolic classification. This is the actual basis that permits us to ensure consistency in the quantitative weighting of spatial relationships and to deal with extended entities beyond the limits of the minimum embedding rectangle approximation.

A prototype retrieval engine is described, and experimental results are reported that indicate the performance of the proposed model with respect to a representation based on a global color histogram, and to a representation that uses centroids orientation to model spatial relationships between color clusters.

The rest of the chapter is organized into five sections and a conclusion. First, to evidence the innovative aspects of weighted walkthroughs, in the remainder of this section, we discuss previous work on modeling techniques for representation and comparison of spatial relationships as developed in the context of image databases (Section 10.1.1).

In Section 10.2, we introduce the representation of chromatic spatial arrangement based on color clusters and their mutual spatial relationships. In particular, we define weighted walkthroughs as original techniques for modeling spatial relationships and discuss their theoretical foundations and properties. Efficient derivation of weighted walkthroughs is expounded in Section 10.3. In Section 10.4, the image representation is cast to a graph theoretical model, and a graph matching approach for the efficient computation of image similarity is prospected. A retrieval engine based on this model is briefly described in Section 10.5. In Section 10.6, we report a two-stage evaluation of the effectiveness of the proposed model, focusing first on a benchmark of basic synthetic arrangements of three colors, and then on a database of real images. Finally, conclusions are drawn in Section 10.7.

10.1.1 Related Work on Modeling Techniques for Representing Spatial Relationships

Several different solutions have been practiced to model spatial relationships in image databases. In particular, at the higher level, representation structures for spatial relationships can be distinguished into object-based and relation-based structures.

The first group comprises those structures that treat spatial relationships and visual information as one inseparable entity. In these approaches, spatial relationships are not explicitly stored, but visual information is included in the representation. As a consequence, spatial relationships are retrieved examining objects coordinates. Object-based structures are based on space partitioning techniques that allow a spatial entity to be located in the space that it occupies. In that some of the data structures used for the indexing of *n*-dimensional points can also handle, in addition to points, spatial objects such as rectangles, polygons, or other geometric bodies, they are particularly suited to being employed as spatial access methods to localize spatial entities in an image. According to this, object-based representations rely on *R*-trees [59], R^+ [60], R^* [61], and their variations [62], [63]. *R*-trees are commonly used to represent the spatial arrangement of rectangular regions and are probably the most popular spatial representation, due to their easy implementation. *R*-trees are particularly effective for searching points or regions. R^+ and R^* trees are improvements of the *R*-tree based on different philosophies. Applications exploiting the spatial properties of these representations have been used mainly in the context of geographical information systems (GISs).

Structures in the second category do not include visual information and preserve only a set of spatial relationships, discarding all uninteresting relationships. Objects are represented symbolically, and spatial relationships explicitly. These approaches may address topological set-theoretical concepts (e.g., inclusion, adjacency, or distance) [64], [65] or directional constructs (e.g., above or below) [19], [66], [67], [68]. In both cases, relationships can be interpreted over a finite set of predefined (symbolic) classes [65], [66], or they can be associated with numeric descriptors taking values in dense spaces [19], [64]. The latter approach enables the use of distance functions that change with continuity and avoid classification thresholds, thus making them better able to cope with the requirements of retrieval by visual similarity [21], [69].

In Reference [64], the topological relationship between pixel sets is encoded by the emptiness/nonemptiness of the intersections between their inner, border, and outer parts. In Reference [65], this approach is extended to the nine-intersection model, so as to allow the representation of topological relationships between regions, lines, and points. Each object is represented as a point set with an interior region, a boundary, and an exterior region. The topological relationship between two objects is described considering the nine possible intersections of their interior, boundary, and exterior.

In References [67], [68], [70], the directional relationship between point-like objects is encoded in terms of the relative quadrants in which the two points are lying. Directional relationships are usually the strict relationships north, south, east, and west. To these are often added the mixed directional relationships northeast, northwest, southeast, and southwest. Other solutions consider the positional directional relationships: left, right, above, and below. Developing on this model, directional spatial relationships are extended to the case of points and lines in Reference [71], while in Reference [72], direction relations for crisp regions are proposed. Generalization of the directional model to regions with broad boundaries is considered in Reference [73].

In the theory of symbolic projection, which underlies a large part of the literature on image retrieval by spatial similarity, both directional and topological relationships between the entities in a two-dimensional (2-D) scene are reduced to the composition of the qualitative ordering relationships among their projections on two reference axes [27], [66]. In the original formulation [66], spatial entities are assimilated to points (usually the centroids) to avoid overlapping and to ensure a total and transitive ordering of the projections on each axis. This permits the encoding of the bidimensional arrangement of a set of entities into a sequential structure, the 2-D-string, which reduces matching from quadratic to linear complexity. However, this point-like representation loses soundness when entities have a complex shape or when their mutual distances are small with respect to individual dimensions. Much work has been done around this model to account for the extent of spatial entities, trading the efficiency of match for the sake of representation soundness. In the 2-DG-string and the 2-DC-string, entities are cut into subparts with disjoint convex hulls [74], [75]. In the 2-D-B string [76], [77], the mutual arrangement of spatial entities is represented in terms of the interval ordering of the projections on two reference axes. Because projections on different axes are independent, the representation subtends the assimilation of objects to their minimum embedding rectangles, which largely reduces the capability to discriminate perceptually distant arrangements. In References [78] and [79], this limit is partially smoothed by replacing extended objects through a finite set of representative points. In particular, in Reference [78], the directional relationship between two entities is interpreted as the union of the primitive directions (up, up-right, right, down-right, down, down-left, left, up-left, coincident), capturing the displacement between any of their respective representative points.

In general, the effectiveness of qualitative models is basically limited by inherent Boolean classification thresholds that determine discontinuities between perceived spatial arrangements and their representation. This hurdles the establishment of quantitative metrics of similarity and basically limits the robustness of comparison. These limits of Boolean matching are faced in quantitative models by associating spatial relationships with numeric values, which enables the evaluation of a continuous distance between nonexact matching arrangements. In the most common approach, directional information is represented through the orientation of the line connecting object centroids [19], [80]. This type of representation inherently requires that extended entities be replaced by a single representative point used to take the measure of orientation. This still limits the capability to distinguish perceptually dissimilar configurations. Representations based on directional histograms have partially solved this limit [81], [82], [83]. The approach in Reference [81] avoids assimilating an object to representative points, like the centroid, or to the minimum bounding rectangle, by computing the histogram of angles between any two points in both the objects. This histogram, normalized by the maximum frequency, represents the directional relationship between the two objects. In Reference [82], histograms are extended to consider pairs of longitudinal sections instead of pairs of points. In this way, it is possible to exploit the power of integral calculus to ensure the processing of raster data as well as of vector data, explicitly considering both angular and metric information. Instead, in Reference [83], the histogram of angles is modified by incorporating both angles and labeled distances. The set of angles from any pixel on the boundaries of two spatial entities expresses their directional relationships. In summary, these angle histogram approaches provide quantitative representation of directional relationships, but they do not provide explicit metric (distance) information and do not support the extraction of topological spatial relationships like "inside" or "overlap."

10.2 Modeling Spatial Arrangements of Color

Using a clustering process [84], the color histogram of an image can be partitioned into a few cohesive clusters, which can be effectively represented by their average color without significant loss of information for the evaluation of similarity. In general, using the CIE $L^*u^*v^*$ color space for color representation, we found that a number of clusters not higher than 8, at most 16, is definitely sufficient to maintain a nonambiguous association between an image and its reduced representation, in which colors are replaced by the average value of their cluster. However, in the backprojection from the color space to the image, each color cluster may be split into several nonmutually connected image segments. This produces an exceedingly complex model, that does not reflect the human capability to merge multiple regions with common chromatic attributes (see Figure 10.1a).

To overcome the limitation, we consider the pixels of each color cluster as a single spatial entity, regardless of their spatial distribution and of their connection in the image space (see Figure 10.1b). The entity is associated with the triple of $L^*u^*v^*$ normalized coordinates of the average color in the cluster. Clusters are also associated with their number of pixels, even if this has only a limited significance due to the capability of the clustering algorithm to produce sets with an approximately equal number of pixels.

10.2.1 Representing Spatial Relationships between Color Clusters

The spatial layout of color clusters is usually complex: color clusters are usually not connected; their mutual distances may be small with respect to their dimensions; and they may be tangled in a complex arrangement evading any crisp classification. These

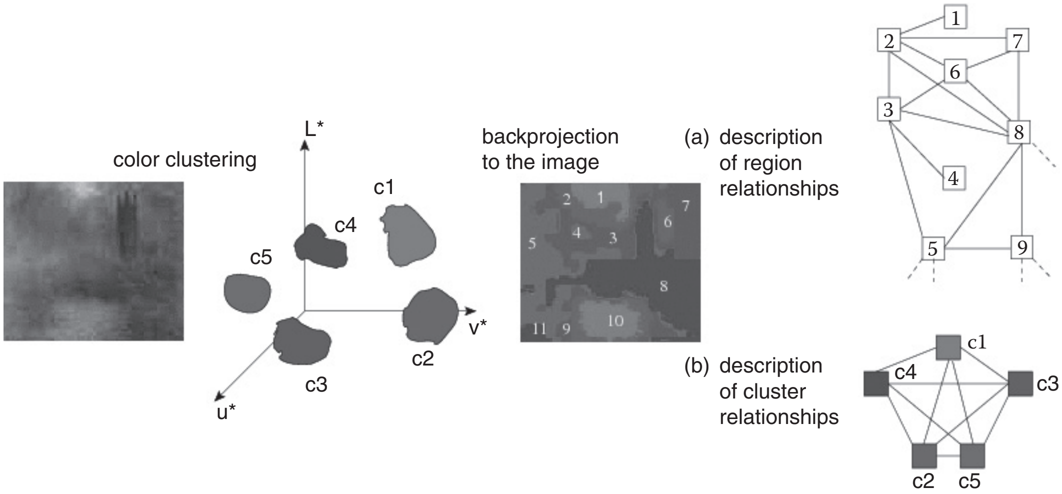


FIGURE 10.1
Pixels are grouped in the color space by using chromatic similarity, so that image content is effectively partitioned into a few clusters. (a) Backprojection in the image space results in a high number of separated segments, yielding an exceedingly complex model for the image. (b) All the pixels obtained from the backprojection of a common cluster are collected within a single entity in the image space.

complexities cannot be effectively managed using conventional spatial descriptors based on centroids or embedding rectangles. To overcome the limit, spatial relationships between color clusters are represented with weighted walkthroughs [21]. In the rest of this subsection, we further develop the original model of weighted walkthroughs to fit the requirements of retrieval by color similarity.

10.2.1.1 Weighted Walkthroughs

In a Cartesian reference system, a point a partitions the plane into four quadrants: upper-left, upper-right, lower-left, and lower-right. These quadrants can be encoded by an index pair $< i, j >$, with i and j taking values ±1. In this perspective, the directional relationship between the point a and an extended set B, can be represented by the number of points of B that are located in each of the four quadrants. This results in four weights $w_{\pm 1, \pm 1}(a, B)$ that can be computed with an integral measure on the set of points of B:

$$w_{i,j}(a, B) = \frac{1}{|B|} \int_B C_i(x_b - x_a) C_j(y_b - y_a)\, dx_b dy_b \tag{10.1}$$

where $|B|$ denotes the area of B and acts as dimensional normalization factor $\langle x_a, y_a \rangle$ and $\langle x_b, y_b \rangle$, respectively, denote the coordinates of the point a, and of points $b \in B$ (see Figure 10.2). The terms $C_{\pm 1}(\cdot)$ denote the characteristic functions of the positive and negative real semi-axes $(0, +\infty)$ and $(-\infty, 0)$, respectively. In particular, $C_{\pm 1}(t)$ are defined in the following way:

$$C_{-1}(t) = \begin{cases} 1 & \text{if } t < 0 \\ 0 & \text{otherwise} \end{cases} \qquad C_1(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{10.2}$$

where, according to Equation 10.1, $t = x_b - x_a$ for $C_i(\cdot)$, and $t = y_b - y_a$ for $C_j(\cdot)$.

The model can be directly extended to represent the directional relationship between two extended sets of points A and B, by averaging the relationships between the individual points of A and B:

$$w_{i,j}(A, B) = \frac{1}{|A||B|} \int_A \int_B C_i(x_b - x_a) C_j(y_b - y_a)\, dx_b dy_b dx_a dy_a \tag{10.3}$$

In doing so, the four-tuple $w(A, B)$ provides a measure of the number of pairs of points in A and B that have a displacement that falls within each of the four directional relationships: $w_{1,1}$ evaluates the number of point pairs $a \in A$ and $b \in B$ such that b is upper-right from a;

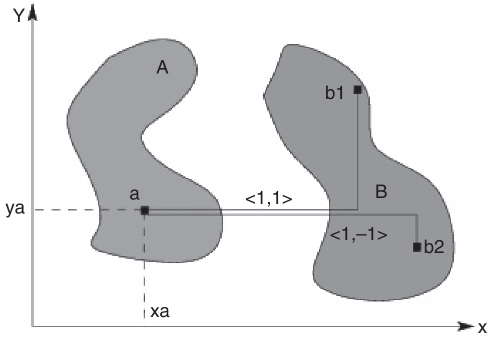


FIGURE 10.2
Walkthroughs connecting the point $a \in A$, to two points $b_1, b_2 \in B$. Because b_1 is in the upper-right quadrant of a, it contributes to the weight $w_{1,1}$. Being b_2 in the lower-left quadrant of a, it contributes to the weight $w_{1,-1}$.

in a similar manner, $w_{-1,1}$ evaluates the number of point pairs such that b is upper-left from a; $w_{1,-1}$ evaluates the number of point pairs such that b is lower-right from a; and $w_{-1,-1}$ evaluates the number of point pairs such that b is lower-left from a.

10.2.1.2 Properties of Weighted Walkthroughs

Because the functions $C_{\pm 1}(.)$ are positive defined, and due to the normalization factor in Equation 10.3, the four weights $w_{i,j}$ are adimensional positive numbers. They are also antisymmetric, that is, $w_{i,j}(A, B) = w_{-i,-j}(B, A)$:

$$\int_A \int_B C_i(x_b - x_a)C_j(y_b - y_a)\, dx_b dy_b\, dx_a dy_a = \int_A \int_B C_{-i}(x_a - x_b)C_{-j}(y_a - y_b)\, dx_b dy_b\, dx_a dy_a$$

as a direct consequence of Equation 10.3, and the antisymmetric property of characteristic functions (i.e., $C_{\pm 1}(t) = C_{\mp 1}(-t)$).

In addition, weighted walkthroughs between two sets A and B are invariant with respect to shifting and zooming of the two sets:

$$w_{i,j}(\alpha A + \beta, \alpha B + \beta) = w_{i,j}(A, B)$$

Shift invariance descends from the fact that $w_{i,j}(A, B)$ is a relative measure (i.e., it depends on the displacement between points in A and B rather than on their absolute position). Scale invariance derives from integration linearity and from the scale invariance of characteristic functions $C_{\pm 1}(\cdot)$.

More importantly, weights inherit from the integral operator of Equation 10.3 a major property of compositionality, by which the weights between A and the union $B_1 \cup B_2$ can be derived by linear combination of the weights between A and B_1, and between A and B_2:

THEOREM 10.2.1
For any point set A, and for any two disjoint point sets B_1 and B_2 (i.e., $B_1 \cap B_2 = \oslash$, and $B_1 \cup B_2 = B$):

$$w_{i,j}(A, B) = w_{i,j}(A, B_1 \cup B_2) = \frac{|B_1|}{|B_1 \cup B_2|} w_{i,j}(A, B_1) + \frac{|B_2|}{|B_1 \cup B_2|} w_{i,j}(A, B_2) \qquad (10.4)$$

PROOF From Equation 10.3 directly descends

$$\begin{aligned} w_{i,j}(A, B_1 \cup B_2) \cdot |A| \cdot |B_1 \cup B_2| &= \int_A \int_{B_1 \cup B_2} C_i(x_b - x_a)C_j(y_b - y_a) dx_b dy_b dx_a dy_a \\ &= \int_A \int_{B_1} C_i(x_b - x_a)C_j(y_b - y_a) dx_b dy_b dx_a dy_a \\ &\quad + \int_A \int_{B_2} C_i(x_b - x_a)C_j(y_b - y_a) dx_b dy_b dx_a dy_a \\ &= w_{i,j}(A, B_1) \cdot |A| \cdot |B_1| + w_{i,j}(A, B_2) \cdot |A| \cdot |B_2| \end{aligned}$$

and dividing both sides by the term $|A| \cdot |B_1 \cup B_2|$, the thesis of the theorem follows. ■

The property of compositionality permits the derivation of the four-dimensional integral of Equation 10.3 through the linear combination of a number of terms corresponding to subintegrals taken over elementary domains for which the weights can be easily computed in closed form. In particular, the four-tuple of weighted walkthroughs can be easily computed over rectangular domains. For example, the weight $w_{1,1}$ between two rectangular

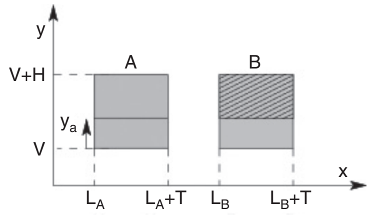


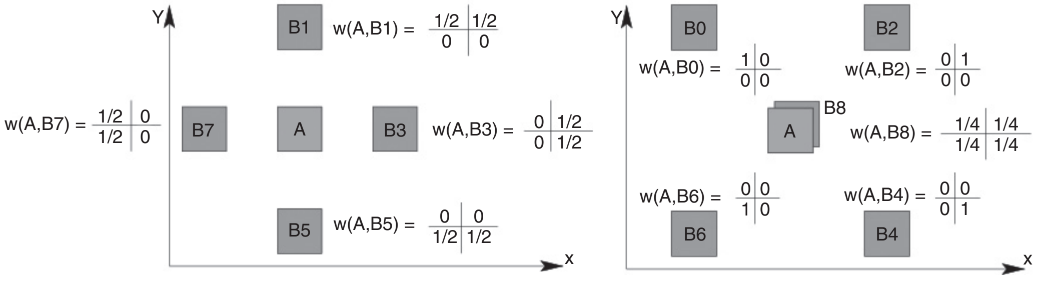
FIGURE 10.3
Determination of $w_{1,1}(A, B)$.

entities with projections that are disjoint along the X axis and perfectly aligned along the Y axis, is computed as follows:

$$
\begin{aligned}
w_{11}(A, B) &= \frac{1}{T^2H^2}\int_A\int_B C_1(x_b - x_a)C_1(y_b - y_a)dx_b dy_b dx_a dy_a \\
&= \frac{1}{T^2H^2}\int_{L_A}^{L_A+T} dx_a \int_{L_B}^{L_B+T} dx_b \int_V^{V+H}\int_{y_a}^{V+H} dy_b dy_a \\
&= \frac{T^2}{T^2H^2}\int_V^{V+H}\left[V + H - y_a\right]dy_a = \frac{1}{H^2}\left[(V+H)y_a - \frac{y_a^2}{2}\right]_V^{V+H} = \frac{1}{2}
\end{aligned}
$$

where, as shown in Figure 10.3, the integration domain along the y dimension of B is limited to the set of points such that $yb > ya, \forall y_a \in A$. Similar computations permit the derivation of the weights $w_{i,j}$ among rectangular domains arranged in the nine basic cases (Figure 10.4a and Figure 10.4b) that represent the possible relationships occurring between two elementary rectangles. This has particular relevance in the context of a digital image with a discrete domain, constituted by individual pixels, that can be regarded as a grid of elementary rectangular elements. In this way, the discrete case can be managed by using the results derived in the continuous domain for the basic elements.

Based on the property of compositionality, and the existence of a restricted set of arrangements between basic elements, if A and B are approximated by any multirectangular shape (see Figure 10.5a), their relationship can be computed by exploiting Equation 10.4 on rectangular domains. According to this, the property of compositionality is used in the computation of weighted walkthroughs between two color regions A and B (see



(a) (b)

FIGURE 10.4
The tuples of weights for the nine basic arrangements between rectangular entities. The weights are represented as elements of a two-by-two matrix. (a) Projections of two rectangular entities are aligned along one of the coordinate axes; (b) disjoint projections and perfect overlap of rectangular entities.

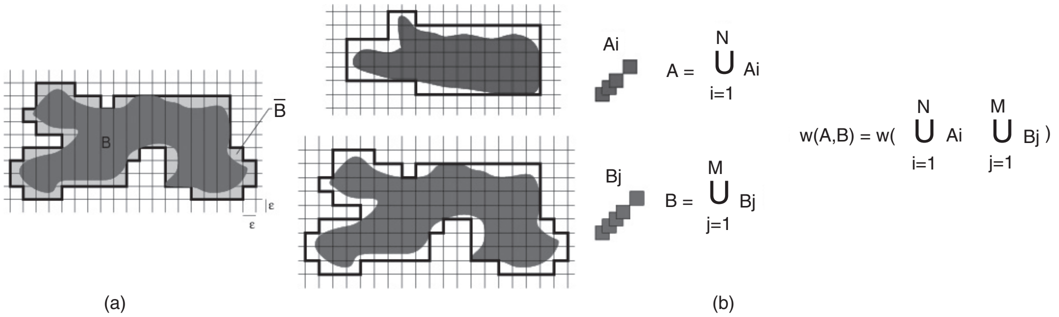


FIGURE 10.5
(a) Entity B is approximated by the minimum embedding multirectangle $\bar{B}$ made up of elements of size $\epsilon \times \epsilon$; (b) Computation of weighted walkthroughs between A and B is reduced to that of a set of rectangular parts arranged in the nine reference positions.

Figure 10.5b), as well as in the composition of relationships between multiple regions within the same color cluster (Figure 10.6).

Finally, it can be observed that the sum of the four weights is equal to one in each of the nine basic cases. As a consequence, the four weights undergo to the following bound:

THEOREM 10.2.2
For any two multirectangular pixel sets A and B, the sum of the four weights is equal to 1:

$$\sum_{i=\pm 1} \sum_{j=\pm 1} w_{i,j}(A, B) = 1 \tag{10.5}$$

PROOF Demonstration runs by induction on the set of rectangles that composes A and B. By the property of compositionality (Theorem 10.2.1), for any partition of B in two disjoint subparts B_1 and B_2, the coefficients of $w(A, B)$ can be expressed as

$$w_{i,j}(A, B) = \frac{|B_1|}{|B_1 \cup B_2|} w_{i,j}(A, B_1) + \frac{|B_2|}{|B_1 \cup B_2|} w_{i,j}(A, B_2)$$

Because this is a convex combination, that is,

$$\frac{|B_1|}{|B_1 \cup B_2|} + \frac{|B_2|}{|B_1 \cup B_2|} = 1$$

coefficients of $w(A, B)$ are a convex combination of coefficients of $w(A, B_1)$ and $w(A, B_2)$, respectively, and so is also the sum of the coefficients themselves. This implies that, by

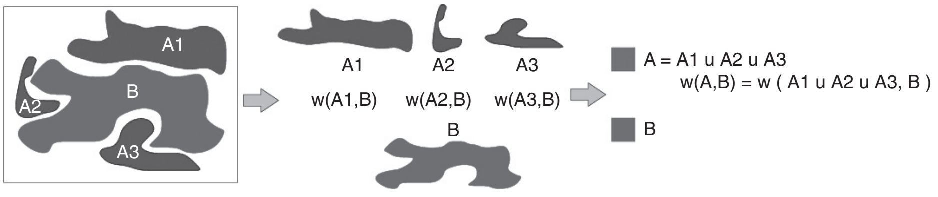


FIGURE 10.6
Property of compositionality applied to the relationship between the nonconnected color cluster A (composed of three segments) and the color cluster B (composed of one segment).

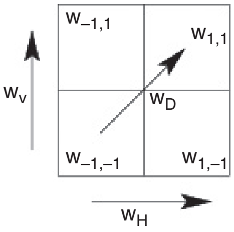


FIGURE 10.7
Directional indices.

recursive decomposition of A and B, the sum of the coefficients of $w(A, B)$ is a convex combination of the sum of the coefficients of the weighted walkthroughs between elementary rectangles mutually arranged in the basic cases.

In all the basic cases of Figure 10.4a and Figure 10.4b, the sum of weights is equal to 1. This implies that any convex combination of the sum of the coefficients of the weighted walkthroughs among any set of elementary rectangles mutually arranged in the basic cases is equal to 1. ■

10.2.1.3 Distance between Weighted Walkthroughs

Because the four weights have a sum equal to 1, they can be replaced, without loss of information, with three directional indexes, taking values within 0 and 1 (Figure 10.7):

$$\begin{aligned} w_H(A, B) &= w_{1,1}(A, B) + w_{1,-1}(A, B) \\ w_V(A, B) &= w_{-1,1}(A, B) + w_{1,1}(A, B) \\ w_D(A, B) &= w_{-1,-1}(A, B) + w_{1,1}(A, B) \end{aligned} \tag{10.6}$$

In doing so, $w_H(A, B)$ and $w_V(A, B)$ account for the degree by which B is on the right, and up of A, respectively; $w_D(A, B)$ accounts for the degree by which A and B are aligned along the diagonal direction of the Cartesian reference system.

In order to compare spatial arrangements occurring between two pairs of spatial entities, the three directional indexes are used. In particular, we experimentally found that a city-block distance constructed on the indices provides effective results in terms of discrimination accuracy between different arrangements. According to this, composition of differences in homologous directional indexes is used as the metric of dissimilarity $\mathcal{D}_S$ for the relationships between two pairs of entities $\langle A, B\rangle$, and $\langle \bar{A}, \bar{B}\rangle$ represented by weights tuples w and $\bar{w}$:

$$\begin{aligned} \mathcal{D}_S(w, \bar{w}) &= \alpha_H \cdot |w_H - \bar{w}_H| + \alpha_V \cdot |w_V - \bar{w}_V| + \alpha_D \cdot |w_D - \bar{w}_D| \\ &= \alpha_H \cdot d_H(w, \bar{w}) + \alpha_V \cdot d_V(w, \bar{w}) + \alpha_D \cdot d_D(w, \bar{w}) \end{aligned} \tag{10.7}$$

where α_H, α_V, and α_D are a convex combination (i.e., they are nonnegative numbers with sum equal to 1), and d_H, d_V, and d_D are the distance components evaluated on the three directional indexes of Equation 10.6. In our framework, we experimentally found that better results are achieved by equally weighting the three distance components ($\alpha_H = \alpha_V = \alpha_D = 1/3$). Due to the city-block structure, $\mathcal{D}_S$ is nonnegative ($\mathcal{D}_S \geq 0$), autosimilar ($\mathcal{D}_S(w, \bar{w}) = 0$ iff $w = \bar{w}$), symmetric ($\mathcal{D}_S(w, \bar{w}) = \mathcal{D}_S(\bar{w}, w)$) and triangular (for any three weights tuples $w, \bar{w}, \hat{w}$: $\mathcal{D}_S(w, \bar{w}) \leq \mathcal{D}_S(w, \hat{w}) + \mathcal{D}_S(\hat{w}, \bar{w})$). In addition, $\mathcal{D}_S$ is normal (i.e., $\mathcal{D}_S \in [0, 1]$) as a consequence of the bound existing on the sum of the weights (Theorem 10.2.2). As an example, Figure 10.8 shows the distance computation between two spatial arrangements.

Weights also satisfy a basic property of continuity by which small changes in the shape or arrangement of entities result in small changes of their relationships. This results in the following theorem for the distance between spatial arrangements:

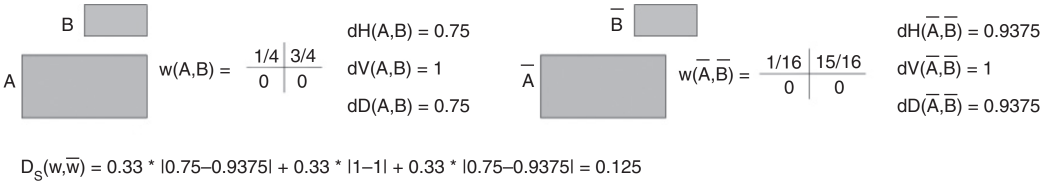


FIGURE 10.8
Spatial distance $\mathcal{D}_S$ between two pairs of entities $\langle A, B\rangle$ and $\langle \bar{A}, \bar{B}\rangle$.

THEOREM 10.2.3
Let A and B be a pair of pixel sets, and let $\bar{B}$ be the minimum multirectangular extension of B on a grid of size ϵ (see Figure 10.5a). Let B_ϵ denote the difference between $\bar{B}$ and B (i.e., $\bar{B} = B \cup B_\epsilon$ and $B \cap B_\epsilon = \oslash$). The distance $\mathcal{D}_S(w(A, B), w(A, \bar{B}))$ between the walkthroughs capturing the relationships between A and B, and between A and $\bar{B}$ undergoes the following bound:

$$\mathcal{D}_S(w(A, B), w(A, \bar{B})) \leq \frac{B_\epsilon}{\bar{B}} \tag{10.8}$$

PROOF Separate bounds are derived for the three distance components d_H, d_V, and d_D. By the property of compositionality (Theorem 10.2.1), $d_H(w(A, B), w(A, \bar{B}))$ can be decomposed as

$$\begin{aligned}
d_H(w(A, B), w(A, \bar{B})) &= |(w_{1,1}(A, B) + w_{1,-1}(A, B)) - (w_{1,1}(A, \bar{B}) + w_{1,-1}(A, \bar{B}))| \\
&= |(w_{1,1}(A, B) + w_{1,-1}(A, B)) - \left(\frac{B}{\bar{B}}(w_{1,1}(A, B) + w_{1,-1}(A, B)\right) \\
&\quad + \frac{B_\epsilon}{\bar{B}}(w_{1,1}(A, B_\epsilon) + w_{1,-1}(A, B_\epsilon)))| \\
&= \frac{B_\epsilon}{\bar{B}}|(w_{1,1}(A, B) + w_{1,-1}(A, B)) - (w_{1,1}(A, B_\epsilon) + w_{1,-1}(A, B_\epsilon))| \\
&= \frac{B_\epsilon}{\bar{B}} d_H(w(A, B), w(A, B_\epsilon))
\end{aligned}$$

which, by the normality of $d_H(\cdot)$, yields

$$d_H(w(A, B), w(A, \bar{B})) \leq \frac{B_\epsilon}{\bar{B}}$$

The same estimate can be applied to $d_V(w(A, B), w(A, \bar{B}))$ and $d_D(w(A, B), w(A, \bar{B}))$, from which the thesis of the theorem follows. ■

10.3 Efficient Computation of Weights

In the straightforward approach, if A and B are decomposed in N and M rectangles, respectively, the four weights of their directional relationship can be computed by repetitive composition of the relationships between the N parts of A and the M parts of B:

$$w(A, B) = w\left(\bigcup_{n=1}^{N} A_n, \bigcup_{m=1}^{M} B_m\right) = \frac{1}{|A||B|} \sum_{n=1}^{N} |A_n| \sum_{m=1}^{M} |B_m| \cdot w(A_n, B_m) \tag{10.9}$$

If component rectangles of A and B are cells of a regular grid partitioning the entire picture, each elementary term $w(A_n, B_m)$ is one of the four-tuples associated with the nine basic arrangements of Figure 10.4a and Figure 10.4b. This permits the computation of $w(A, B)$ in time $O(N \cdot M)$.

A more elaborate strategy permits the derivation of the relationship with a complexity that is linear in the number of cells contained in the intersection of the bounding rectangles of the two entities. This is expounded in the rest of this section.

10.3.1 Representation of Spatial Entities

We assume that each entity is approximated as a set of rectangular cells taken over a regular grid partitioning the entire picture along the horizontal and vertical directions of the Cartesian reference system. The set of cells comprising each entity is partitioned in any number of segments. Each of these segments is assumed to be connected, but not necessarily maximal with respect to the property of the connection (as an example, in Figure 10.6, the nonconnected entity A is decomposed into the three connected segments A_1, A_2, and A_3). Here, we expound the representation of segments and the computation of their mutual relationships. Relationships between the union of multiple segments are derived by direct application of the property of compositionality (Equation 10.4). The following clauses illustrate the derivation:

$$\begin{array}{lll}
WW^{0,0}_{i,j} & = & 1 \quad \text{if the cell } \langle i, j\rangle \text{ is part of } A \\
 & & 0 \quad \text{otherwise} \\
WW^{-1,0}_{i,j} & = & 0 \quad \text{if } j = 0 \text{ (i.e., } j \text{ is the leftmost column of } A) \\
 & & WW^{-1,0}_{i,j-1} + WW^{0,0}_{i,j} \quad \text{otherwise} \\
WW^{1,0}_{i,j} & = & \text{is derived by scanning the row } i \text{ if } j = 0 \\
 & & WW^{1,0}_{i,j-1} - WW^{0,0}_{i,j} \quad \text{otherwise} \\
WW^{0,-1}_{i,j} & = & 0 \quad \text{if } i = 0 \text{ (i.e., } i \text{ is the lowermost row of } A) \\
 & & WW^{0,-1}_{i-1,j} + WW^{0,0}_{i-1,j} \quad \text{otherwise} \\
WW^{0,1}_{i,j} & = & \text{is derived by scanning the column } j \text{ if } i = 0 \\
 & & WW^{0,1}_{i-1,j} - WW^{0,0}_{i,j} \quad \text{otherwise} \\
WW^{-1,1}_{i,j} & = & 0 \quad \text{if } j = 0 \\
 & & WW^{-1,1}_{i,j-1} + WW^{0,1}_{i,j-1} \quad \text{otherwise} \\
WW^{-1,-1}_{i,j} & = & 0 \quad \text{if } i = 0 \text{ or } j = 0 \\
 & & WW^{-1,-1}_{i,j-1} + WW^{0-1}_{i,j-1} \quad \text{otherwise} \\
WW^{1,-1}_{i,j} & = & 0 \quad \text{if } i = 0 \\
 & & WW^{1,-1}_{i-1,j} + WW^{1,0}_{i-1,j} \quad \text{otherwise} \\
WW^{1,1}_{i,j} & = & N - WW^{0,0}_{i,j} - WW^{0,1}_{i,j} - WW^{1,0}_{i,j} \quad \text{if } j = 0 \text{ and } i = 0 \\
 & & WW^{1,1}_{i,j-1} - WW^{0,0}_{i,j} - WW^{0,1}_{i,j} \quad \text{if } j = 0 \text{ and } i > 0 \\
 & & WW^{1,1}_{i-1,j} - WW^{0,0}_{i,j} - WW^{1,0}_{i,j} \quad \text{if } j > 0
\end{array} \qquad (10.10)$$

Each segment A is represented by a data structure that encompasses the following information: the number of cells of A, and the indexes $\langle i_l, j_l\rangle$ and $\langle i_u, j_r\rangle$ of the cells of the lower-left and of the upper-right corners of the bounding rectangle of A. The segment A is also associated with a matrix WW with size equal to the number of cells in the bounding rectangle of A, which associates each cell $\langle i, j\rangle$ in the bounding rectangle of A with a nine-tuple $WW_{i,j}$ that encodes the number of cells of A in each of the nine directions centered in the cell $\langle i, j\rangle$: $WW^{0,0}_{i,j}$ is equal to 1 if the cell $\langle i, j\rangle$ is part of A, and it is equal to zero otherwise; $WW^{1,0}_{i,j}$ is the number of cells of A that are on the right of cell $\langle i, j\rangle$ (i.e., the number of cells of A with indexes $\langle i, k\rangle$ such that $k > j$); in a similar manner, $WW^{-1,0}_{i,j}$ is the number of cells of A

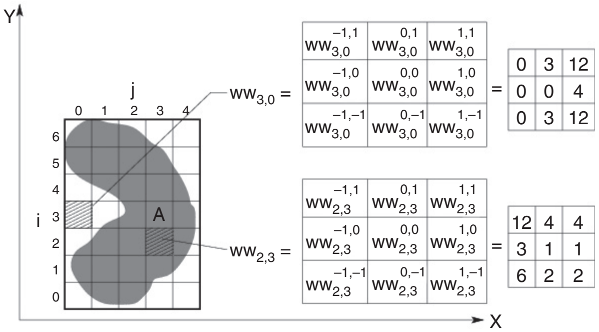

Y
X
j
0 1 2 3 4
i
6 5 4 3 2 1 0
A
$ww_{3,0} =$
$ww_{3,0}^{-1,1}$ $ww_{3,0}^{0,1}$ $ww_{3,0}^{1,1}$
$ww_{3,0}^{-1,0}$ $ww_{3,0}^{0,0}$ $ww_{3,0}^{1,0}$
$ww_{3,0}^{-1,-1}$ $ww_{3,0}^{0,-1}$ $ww_{3,0}^{1,-1}$
=
0 3 12
0 0 4
0 3 12
$ww_{2,3} =$
$ww_{2,3}^{-1,1}$ $ww_{2,3}^{0,1}$ $ww_{2,3}^{1,1}$
$ww_{2,3}^{-1,0}$ $ww_{2,3}^{0,0}$ $ww_{2,3}^{1,0}$
$ww_{2,3}^{-1,-1}$ $ww_{2,3}^{0,-1}$ $ww_{2,3}^{1,-1}$
=
12 4 4
3 1 1
6 2 2

FIGURE 10.9
Examples of the data structure WW, computed for the cells $\langle 3, 0\rangle$ and $\langle 2, 3\rangle$ of the bounding rectangle enclosing the entity A.

that are on the left of $\langle i, j\rangle$, while $WW_{i,j}^{0,1}$ and $WW_{i,j}^{0,-1}$ are the number of cells of A over and below cell $\langle i, j\rangle$, respectively; finally, $WW_{i,j}^{1,1}$ is the number of cells of A that are upper-right from $\langle i, j\rangle$ (i.e., the cells of A with indexes $\langle h, k\rangle$ such that $h > i$ and $k > j$). In a similar manner, $WW_{i,j}^{1,-1}$, $WW_{i,j}^{-1,-1}$ and $WW_{i,j}^{-1,1}$ are the numbers of cells of A that are lower-right, lower-left, and upper-left from the cell $\langle i, j\rangle$, respectively. Figure 10.9 shows the WW matrix computed for two cells of the bounding rectangle of an entity A.

The matrix WW of the segment A is derived in linear time with respect to the number of cells in the bounding rectangle of A. To this end, the elements of the matrix are computed starting from the lower-left corner, covering the matrix by rows and columns. In doing so, the nine coefficients associated with any cell $\langle i, j\rangle$ can be derived by relying on the coefficients of the cells $\langle i - 1, j\rangle$ (lower adjacent), and $\langle i, j - 1\rangle$ (left adjacent).

In the overall derivation, a constant time $O(1)$ is spent for evaluating coefficients of each cell; thus requiring a total cost $O(L_A \cdot H_A)$, where L_A and H_A are the numbers of columns and rows of the bounding box of A, respectively. In addition, the entire column of each cell in the first row, and the entire row of each cell in the first column must be scanned, with a total cost $O(2 \cdot L_A \cdot H_A)$. According to this, the total complexity for the derivation of the overall matrix WW is linear in the number of cells in the bounding rectangle of A.

10.3.2 Computation of the Four Weights

Given two segments A and B, the four weights of their relationship are computed from the respective descriptions, in a way that depends on the intersection between the projections of A and B on the Cartesian reference axes:

- If the projections of A and B have null intersections on both the axes, then the descriptor has only a nonnull weight (and this weight is equal to 1) that is derived in constant time (see Figure 10.4b).
- If the projections of A and B on the Y axis have a nonnull intersection, but the projections on the X axis are disjoints (see, for example, Figure 10.10), then the descriptor has two null elements and is determined with complexity $O(H_{AB})$, where H_{AB} is the number of cells by which the projections intersect along the Y axis. Of course, the complementary case that the projections of A and B have nonnull intersection along the X axis is managed in the same manner.

  We expound here the method for the case in which B is on the right of A (see Figure 10.10). In the complementary case (B on the left of A), the same algorithm serves to derive the relationship $w(B, A)$, which can then be transformed into

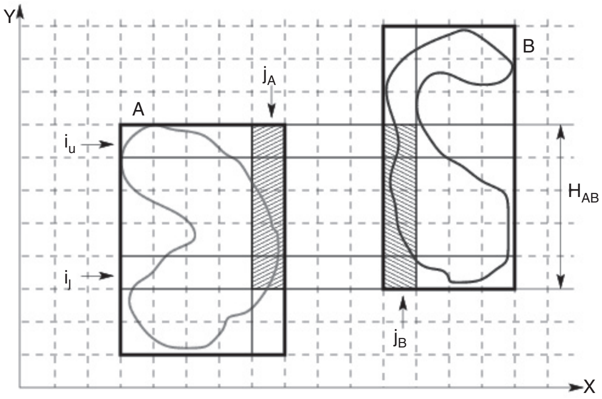


FIGURE 10.10
Projections of bounding rectangles A and B intersect along the Y axis. The gray patterns indicate cells that are scanned in the evaluation of coefficient $w_{1,1}(A, B)$. This is sufficient to evaluate the complete relationship between entities represented by segments A and B.

$w(A, B)$ by applying the property of antisymmetry of weighted walkthroughs. Because all the cells of A are on the left of B, the two upper-left and lower-left weights $w_{-1,1}(A, B)$ and $w_{-1,-1}(A, B)$ are equal to 0. In addition, because the sum of the four weights is equal to 1, the derivation of the upper-right weight $w_{1,1}(A, B)$ is sufficient to fully determine the descriptor (as $w_{1,-1}(A, B) = 1 - w_{1,1}(A, B)$).

The upper-right weight $w_{1,1}(A, B)$ is computed by summing up the number of cells of A that are lower-left or left from cells of B. According to the forms computed in the nine basic cases of Figure 10.4a and Figure 10.4b, for any cell $\langle i, j\rangle$ in A, the contribution to $w_{1,1}(A, B)$ is equal to 1 for each cell of B having indexes $\langle h, k\rangle$ with $h > i$ and $k > j$, and it is equal to 1/2 for each cell of B having indexes $\langle h, k\rangle$ with $h = i$ and $k > j$. In the end of the computation, the total sum is normalized by dividing it by the product of the number of cells in A and B.

By relaying on matrixes WW in the representation of segments A and B, the computation can be accomplished by scanning only once a part of the rightmost column of the bounding box of A and of the leftmost column of the bounding box of B, without covering the entire set of cells in A and B. The algorithm is reported in Figure 10.11. UR denotes the weight $w_{1,1}(A, B)$ being computed. For the simplicity of notation, matrixes WW of segments A and B are denoted by A and B. Notations j_A and j_B denote the indexes of the right column of the bounding box of A and of the left column of the bounding box of B, respectively. Finally, i_l and i_u indicate the indexes of the lowest and topmost rows that contain cells of both A and B, respectively (see Figure 10.10).

In the statement on line 1, the term $(A^{-1,-1}_{i_l,j_A} + A^{0,-1}_{i_l,j_A})$ evaluates the number of cells of A that are lower-left, or lower-aligned with respect to i_l, j_A; for each of these cells, there are no cells of B that are aligned on the right-hand side, and the number of cells of B that are in the upper-right position is equal to the term $(B^{0,0}_{i_l,j_B} + B^{0,1}_{i_l,j_B} + B^{1,0}_{i_l,j_B} + B^{1,1}_{i_l,j_B})$. According to this, statement 1 initializes UR by accounting for the contribution of all the (possibly existing) rows of A that are below row i_l. The statement in line 2 controls a loop that scans the cells in the right column of A and in the left column of B, throughout the height of the intersection of the projections of A and B on the vertical axis. Note that, because i_u is the topmost row of A or of B, there cannot be any other cell of A that is over row

```
1. $UR = (A^{-1,-1}_{i_l,j_A} + A^{0,-1}_{i_l,j_A}) \cdot (B^{0,0}_{i_l,j_B} + B^{0,1}_{i_l,j_B} + B^{1,0}_{i_l,j_B} + B^{1,1}_{i_l,j_B}) \cdot 1;$
2. for $i = i_l : i_u$
3.     $UR = UR + (A^{-1,0}_{i,j_A} + A^{0,0}_{i,j_A}) \cdot ((B^{0,0}_{i,j_B} + B^{1,0}_{i,j_B}) \cdot 1/2 + (B^{0,1}_{i,j_B} + B^{1,1}_{i,j_B}) \cdot 1);$
4. $UR = UR/(N \cdot M);$
```

FIGURE 10.11
Algorithm for the case in which A and B have a null intersection along the X axis.

i_u, and that has any cell of B up-right or aligned-right. The statement 3, in the body of the loop, adds to *UR* the contribution of all the cells of A belonging to row i: $(A^{-1,0}_{i,j_A} + A^{0,0}_{i,j_A})$ is the number of cells of A in row i; each of these cells has $(B^{0,0}_{i,j_B} + B^{1,0}_{i,j_B})$ cells of B aligned on the right-hand side (contributing the weight 1/2), and $(B^{0,1}_{i,j_B} + B^{1,1}_{i,j_B})$ cells of B that are up-right (each contributing the weight 1). The statement in line 4 normalizes the weight.

- When projections of A and B have a nonnull intersection on both the axes (i.e., when the bounding boxes of A and B overlap [see Figure 10.13], all four weights can be different than 0, and three of them must be computed (the fourth can be determined as the complement to 1). The derivation of each of the three weights is accomplished in time linear with respect to the number of cells falling within the intersection of bounding boxes of A and B.

PE: Note that Figure 10.13 is called out before Figure 10.12

We expound here the derivation of $w_{1,1}(A, B)$. Of course, any of the other three weights can be derived in a similar manner, with the same complexity.

The derivation of $w_{1,1}(A, B)$ consists of evaluating how many cells of A have how many cells of B in the upper-right quadrant, in the upper column, in the right row, or coincident. According to the forms computed in the nine basic arrangements of Figure 10.4a and Figure 10.4b, each cell in the upper-right quadrant provides a contribution equal to 1, each cell in the upper column or in the right row provides a contribution equal to 1/2, and each cell coincident provides a contribution equal to 1/4.

Also, in this case, matrices *WW* associated with A and B permit the evaluation by scanning only once a limited set of cells of A and B. The algorithm is reported in Figure 10.12. In this case, indexes i_l, i_r, j_l, and j_r indicate the lower and upper row, and the left and right column of the intersection of bounding boxes of A and B, respectively (see Figure 10.13).

```
1. $UR = (A^{-1,-1}_{i_l,j_l}) \cdot (B^{0,0}_{i_l,j_l} + B^{0,1}_{i_l,j_l} + B^{1,0}_{i_l,j_l} + B^{1,1}_{i_l,j_l}) \cdot 1;$
2. for $i = i_l : i_u$
3.     $UR = UR + (A^{-1,0}_{i,j_l}) \cdot ((B^{0,1}_{i,j_l} + B^{1,1}_{i,j_l}) \cdot 1 + (B^{0,0}_{i,j_l} + B^{1,0}_{i,j_l}) \cdot 1/2);$
4. for $j = j_l : j_r$
5.     $UR = UR + (A^{0,-1}_{i_l,j}) \cdot ((B^{1,0}_{i_l,j} + B^{1,1}_{i_l,j}) \cdot 1 + (B^{0,0}_{i_l,j} + B^{0,1}_{i_l,j}) \cdot 1/2);$
6. for $i = i_l : i_u$
7.     for $j = j_l : j_r$
8.         $UR = UR + (A^{0,0}_{i,j}) \cdot ((B^{1,1}_{i,j}) \cdot 1 + (B^{1,0}_{i,j} + B^{0,1}_{i,j}) \cdot 1/2 + (B^{0,0}_{i,j}) \cdot 1/4);$
9. $UR = UR/(N \cdot M);$
```

FIGURE 10.12
Algorithm for the case in which A and B have a nonnull intersection on both the X and Y axes.

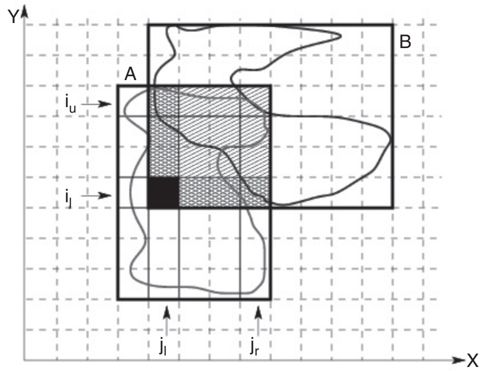


FIGURE 10.13
Projections of bounding rectangles of A and B have a nonnull intersection on both the axes. During the evaluation of relationships, the cells filled with the less dense pattern are scanned once, those with a more dense pattern are scanned twice, and the black cell is scanned three times.

Statement 1 initializes the weight $w_{1,1}(A, B)$, denoted as UR, by summing up the contribution of the $(A^{-1,-1}_{i_l, j_l})$ cells of A that are in the lower-left quadrant of the cell $\langle i_l, j_l\rangle$. The loop in statements 2 and 3 adds to UR the contribution of all the cells of A that are on the left of the intersection area of the bounding boxes of A and B. These cells yield a different contribution on each row i in the range between i_l and i_u. In a similar manner, the loop in statements 4 and 5 adds to UR the contribution of all the cells that are below the intersection area of the bounding boxes of A and B. Finally, the double loop in statements 6, 7, and 8 adds the contribution of the cells of A that fall within the intersection of the bounding boxes of A and B. Statement 9 normalizes the weight.

10.4 Graph Representation and Comparison of Spatial Arrangements

Color clusters and their binary spatial relationships can be suitably represented and compared in a graph–theoretical framework. In this case, an image is represented as an attributed relational graph (ARG):

$$\text{image model} \overset{def}{=} < E, a, w >, \qquad \begin{array}{l} E = \text{set of spatial entities} \\ a : E \rightarrow A \cup \{\text{any}_\text{a}\} \\ w : E \times E \rightarrow W \cup \{\text{any}_\text{s}\} \end{array} \tag{10.11}$$

where spatial entities are represented by vertices in E, and their chromatic features are captured by the attribute label a; spatial relationships are the complete set of pairs in $E \times E$, each labeled by the spatial descriptor w. To accommodate partial knowledge and intentional detail concealment, we also assume that both edges and vertices can take a neutral label any, yielding an exact match in every comparison (i.e., $\forall w \in W$, $\mathcal{D}_S(w, \text{any}_s) = 0$, and $\forall a \in A$, $\mathcal{D}_A(a, \text{any}_a) = 0$).

In so doing, $\mathcal{D}_S$ is the spatial distance defined in Section 10.2.1, while $\mathcal{D}_A$ is the metric of chromatic distance defined in the $L^*u^*v^*$ color space. In particular, the $L^*u^*v^*$ color space has been specifically designed to be "perceptual," this meaning that the distance between two colors with coordinates that are not far apart in the space, can be evaluated by using the Euclidean distance. According to this, attributes a_1 and a_2 of two entities are compared by using an Euclidean metric distance:

$$\mathcal{D}_A(a_1, a_2) \overset{def}{=} \sqrt{\alpha_L \cdot (L^*_{a_1} - L^*_{a_2})^2 + \alpha_u \cdot (u^*_{a_1} - u^*_{a_2})^2 + \alpha_v \cdot (v^*_{a_1} - v^*_{a_2})^2} \tag{10.12}$$

where α_L, α_u, and α_v is a convex combination (i.e., α_L, α_u, and α_v are nonnegative numbers with sum equal to 1). Because there is not a preferred coordinate in the space, we set $\alpha_L, \alpha_u, \alpha_v = 1/3$. Distance $\mathcal{D}_A$ is nonnegative, autosimilar, symmetric, and normal, and satisfies the triangular inequality.

The comparison of the graph models of a query specification $< Q, a^q, w^q >$ and an archive image description $< D, a^d, w^d >$ involves the association of the entities in the query with a subset of the entities in the description. This is represented by an injective function Γ that we call *interpretation*.

The distance between two image models Q and D, under an interpretation Γ can be defined by combining the metrics of chromatic distance $\mathcal{D}_A$, and spatial distance $\mathcal{D}_S$, associated with entity attributes (vertices) and relationship descriptors (edges), respectively. Using an additive composition, this is expressed as follows:

$$\mu^\Gamma(Q, D) \overset{def}{=} \lambda \sum_{k=1}^{N_q} \mathcal{D}_A(q_k, \Gamma(q_k)) + (1 - \lambda) \sum_{k=1}^{N_q} \sum_{h=1}^{k-1} \mathcal{D}_S([q_k, q_h], [\Gamma(q_k), \Gamma(q_h)]) \tag{10.13}$$

where N_q is the number of entities in the query graph Q, and $\lambda \in [0, 1]$ balances the mutual relevance of spatial and chromatic distance: for $\lambda = 1$, distance accounts only for the chromatic component.

In general, given the image models Q and D, a combinatorial number of different interpretations Γ are possible, each scoring a different value of distance. The distance is thus defined as the minimum distance under any possible interpretation:

$$\mu(Q, D) \overset{def}{=} \min_{\Gamma} \mu^\Gamma(Q, D) \tag{10.14}$$

In doing so, computation of the distance between two image models becomes an optimal error-correcting (sub)graph isomorphism problem [85], which is a NP-complete problem with exponential time solution algorithms.

In the proposed application, the problem of matching a query graph Q against a description graph D is faced following the approach proposed in Reference [86]. To avoid exhaustive inspection of all possible interpretations Γ of Q on D, the search is organized in an incremental approach by repeatedly growing a partial assignment of the vertices of the query to the vertices of the description. In so doing, the space of solutions is organized as a tree, where the kth level contains all the partial assignments of the first k entities of the query. Because the function of distance is monotonically growing with the level, any partial interpretation scoring a distance over a predefined threshold of maximum acceptable dissimilarity μ_{max} can be safely discarded without risk of false dismissal. While preserving the exactness of results, this reduces the complexity of enumeration. Following the approach of the A^* algorithm [87], a search is developed in depth-first order by always extending the partial interpretation toward the local optimum, and by backtracking when the scored distance of the current assignment runs over a maximum acceptable threshold. When the inspection reaches a complete interpretation, a match under the threshold is found. This is not guaranteed to be the global optimum, but its scored distance comprises a stricter threshold for acceptable distance that is used to efficiently extend the search until the global optimum is found.

In Reference [86], a look-ahead strategy is proposed that extends the basic A^* schema using an admissible heuristic to augment the cost of the current partial interpretation with a lower estimate of the future cost that will be spent in its extension to a complete match. This permits a more "informed" direction of search and enables the early discard of partial assignments that cannot lead to a final match with acceptable similarity. This reduces the complexity of the search while preserving the optimality of results. The approach results were compatible with the dimension encountered in the application context of retrieval by spatial arrangement.

10.5 A Retrieval System

The metric of similarity in Equation 10.13 and Equation10.14, based on the joint combination of color clusters and weighted walkthroughs, was employed within a prototype retrieval engine.

In the archiving stage, all images are segmented into eight clusters and are represented by eight-vertices complete graphs. This resulted as a trade-off between the accuracy of representation and the efficiency of the graph matching process. The usage of graphs with fixed size permits the exploitation of the metric properties of the graph distance, thus enabling the exploitation of a metric indexing scheme [86]. The system supports two different query modalities: global similarity and *sketch*.

In a query by *global similarity*, the user provides an example by directly selecting an image from the database (see Figure 10.14), and the retrieval engine compares the query

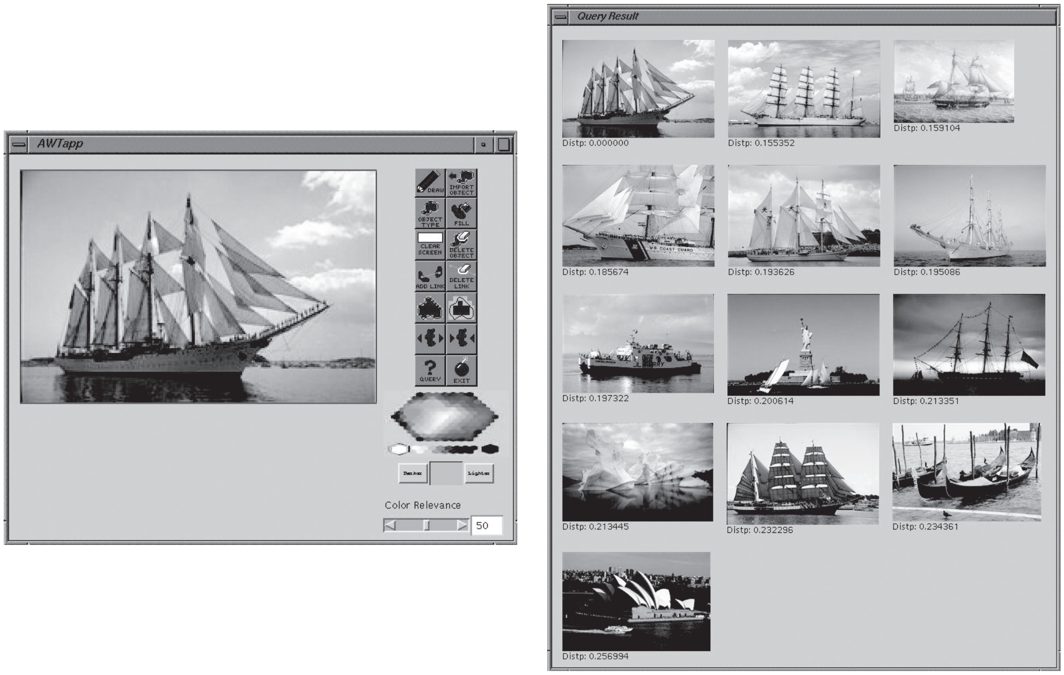


FIGURE 10.14 (See color insert.)
A query by image example (left), and the corresponding retrieval set (right).

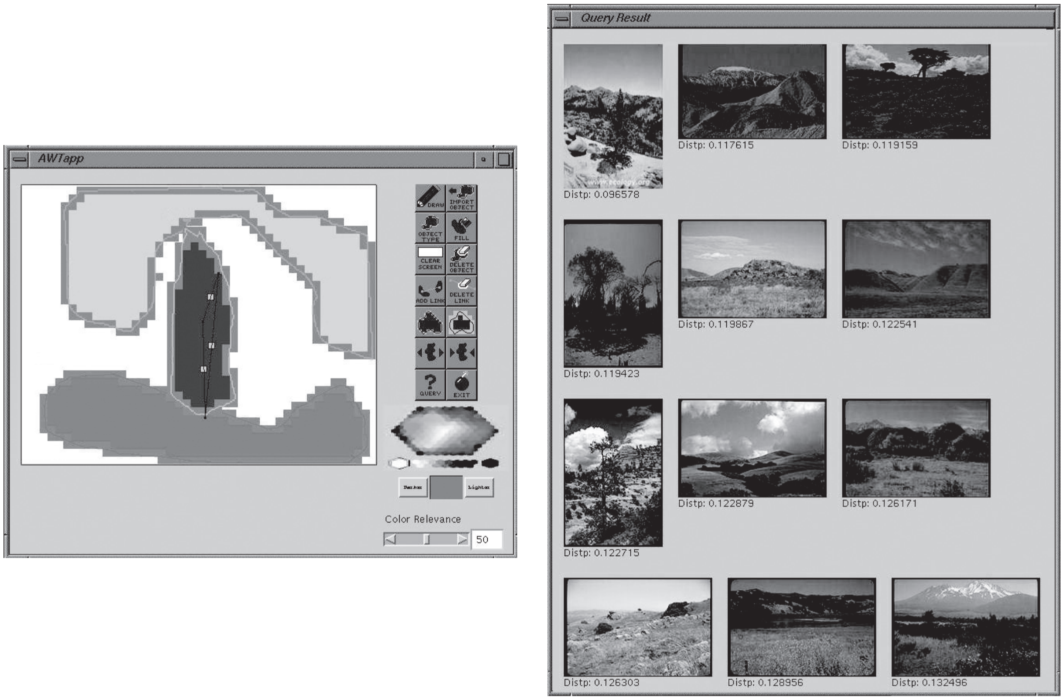


FIGURE 10.15 (See color insert.)
A query by sketch (left), and the corresponding retrieval set (right).

graph with database descriptions. In a query by *sketch*, the user expresses the query by drawing, coloring, and positioning a set of regions that capture only the color patches and relationships that are relevant to the user (see Figure 10.15 and Figure 10.16). From this representation, a query graph is automatically derived following a decomposition approach. Each region corresponds to a variable number of color clusters, depending on its size normalized with respect to that of the drawing area. This has a twofold effect.

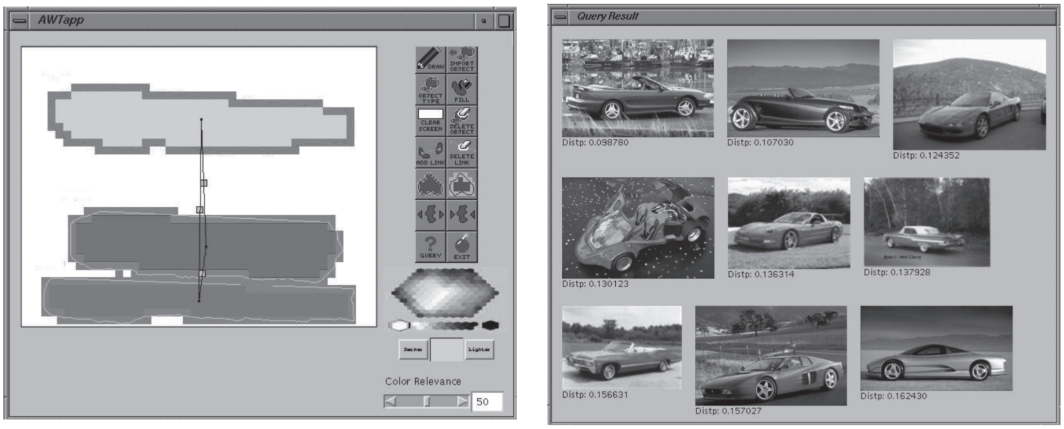


FIGURE 10.16
A query by sketch, and the corresponding retrieval set for λ = 0.5 (color relevance set to 50).

On the one hand, the different relevance of regions, implicitly associated with their size, is considered by splitting them into a different number of graph entities. On the other hand, this partially replicates the behavior of the clustering algorithms, which splits sets of pixels according to their size, thus providing multiple clusters for colors with a predominant occurrence. Relationships between entities are those explicitly drawn by the user. If a region is decomposed in multiple entities, relationships between this region and other regions in the query are extended to all entities derived from the decomposition. The query graph derived from this representation involves a restricted match between the N_q entities in the query against the N_d entities in the database descriptions (with $N_q \leq N_d$).

For both queries, the user is allowed to dynamically set the balance of relevance by which spatial and chromatic distances are combined in the searching process. In the framework of Section 10.4, this is obtained by setting parameter λ in Equation 10.13.

10.5.1 Retrieval Examples

Results are reported for a database of about 10,000 photographic images collected from the Internet. Figure 10.15 illustrates the querying operation: the user draws a sketch of the contour of characterizing color entities and positions them so as to reproduce the expected arrangement in searched images. The sketch is interpreted by the system as a set of color clusters and their spatial relationships and is checked against the descriptions stored in the archive. Matching images are displayed in a separate window, sorted by decreasing similarity from top to bottom and from left to right (see Figure 10.15). The user can tune, with the slide bar *color relevance*, the balance of color and spatial distances to the overall measure.

In Figure 10.16, a query for a different sketch is shown; the interpretation of the sketch takes into account only those spatial relationships that are marked by the user (made explicit by lines on the screen). In this case, the color relevance is set equal to 50, corresponding to $\lambda = 0.5$, so that color and spatial similarities are given equal weight.

Figure 10.14 shows the expression of a query by example. In this case, one of the database images is used as a query example, and the system searches for the most similar images, using all the color entities and relationships that appear in the query representation. In the particular example, the system retrieves the query image in the first position, and other images with a similar arrangement. Some of the retrieved images show a lower consistency in the semantics of the imaged scenes but still have a high relevance in terms of chromatic content and spatial arrangement.

10.6 User-Based Assessment

The perceptual significance of the metric of dissimilarity derived through the joint representation of color and spatial content was evaluated in a two-stage test, focusing first on a benchmark of images representing basic synthetic arrangements of three colors, and then on a database of real images.

10.6.1 A Benchmark Database of Basic Spatial Arrangements of Color

The first stage of the evaluation was oriented to investigate the capability of the proposed model to capture differences and similarities in basic spatial arrangements of colors, by abstracting from other relevant features, such as color distribution and size and shape of color patches.

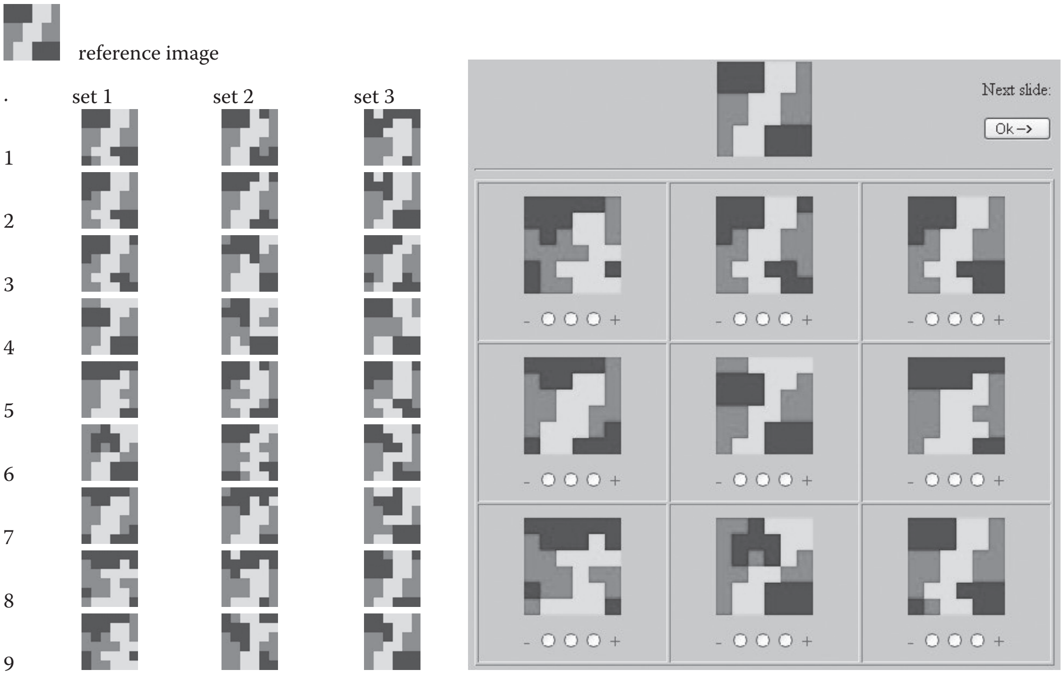


FIGURE 10.17
The columns on the left are the images, listed from top to bottom in increasing order of variation, comprised in the three sets of mutations of the reference image on the top left. A page of the user test for the reference image and the mutation set 1 is shown on the right.

To this end, the evaluation was carried out on a benchmark based on an archive with 6 × 3 × 9 synthetic pictures. The archive was derived from six reference pictures, obtained by different compositions of an equal number of red, yellow, and blue squares within a six-by-six grid. Reference pictures (displayed on the top of the plots of Figure 10.18) were created so as to contain five or six separate regions each. Preliminary pilot tests indicated that this number results in a complexity that is sufficient to prevent the user from acquiring an exact memory of the arrangement. Though these images are completely synthetic and not occurring in real application contexts, they are useful in testing the effectiveness of spatial descriptors independently from chromatic components. In fact, their structure allows for an easier evaluation by the users which can focus on spatial arrangements rather than on the semantics or other image features that could bias the results of the evaluation in the case of real images.

PE: Note that Figure 10.18 is called out before Figure 10.17

For each reference picture, three sets of mutations were derived automatically by a random engine changing the arrangement of blocks through shift operations on randomly selected rows or columns. Each set includes nine variations of the reference picture, which attain different levels of mutation by applying a number of shift operations ranging from one to nine. (Figure 10.17 indicates the level of mutation for the nine variations in each of the three sets of a reference picture.) In order to avoid the introduction of a perceivable ordering, mutations were derived independently (i.e., the mutation at level n was obtained through n shifts on the reference picture rather than through one shift on the mutation at level $n-1$). By construction, the mutation algorithm maintains the overall picture histogram and the multirectangular shape of segments, but it largely increases the fragmentation of regions. Preliminary pilot tests with variations including more than eight regions resulted

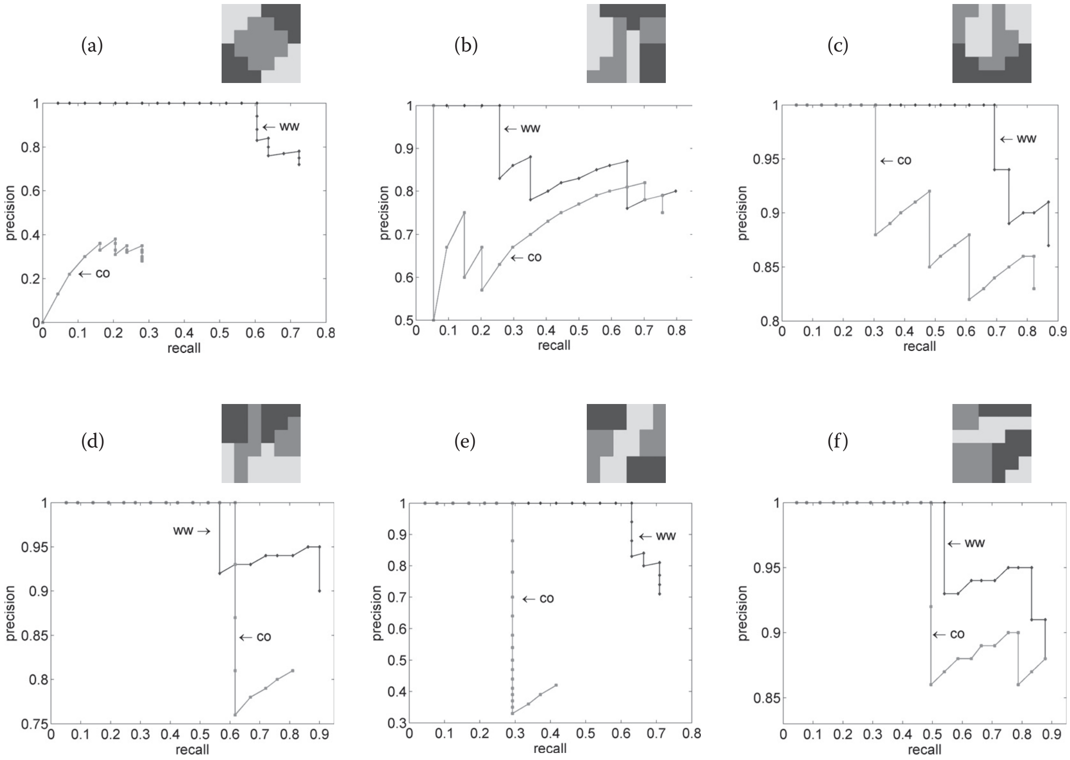


FIGURE 10.18
The six query images used in the test, and their corresponding plots of precision/recall. Plotted values correspond to those obtained by resolving the query using both the weighted walkthroughs (WW), and the centroid orientation (CO).

in major complexity for the user in comparing images and ranking their similarity. The algorithm was thus forced to accept only arrangements resulting in less than eight regions.

The six reference pictures were employed as queries against the 6 × 3 × 9 pictures of the archive, and queries were resolved using the metric of dissimilarity defined in Equation 10.13.

10.6.2 Ground Truth

Evaluation of the effectiveness of retrieval obtained on the benchmark requires a ground-truth about the similarity V_{qd} between each reference picture q and each archive image d. With six queries against an archive of 162 images, this makes 972 values of similarity, which cannot be realistically obtained with a fully user-based rank. To overcome the problem, user rankings have been complemented with inference [88].

Each user was shown a sequence of 3×6 html pages, each showing a reference picture and a set of nine variations. (Figure 10.17 reports a test page, while the overall testing session is available online at http://viplab.dsi.unifi.it/color/test). Pages presenting variations of the same reference picture were presented subsequently, so as to maximize the correlation in the ranking of variations included in different sets. On each page, the user was asked to provide a three-level rank of the similarity between the reference picture and each of the nine variations. To reduce the stress of the test, users were suggested to first search for the most similar images and then extend the rank toward low similarities, thus emphasizing the relevance of high ranks.

The testing session was administered to a sample of 22 volunteers and took an average time ranging between 10 and 21 min, with an average of 14.6 min. This appeared to be a realistic limit for the user capability and willingness in maintaining attention throughout the test. The overall sizing of the evaluation, and, in particular, the number of reference pictures and queries considered, was based on preliminary evaluation of this limit.

User ranks were employed to evaluate a ground value of similarity between each reference picture and its compared variations. In order to reflect the major relevance of higher ranks, a score of three was attributed for each high rank received; a score of 1 was attributed for each intermediate rank. No score was attributed for low ranks, as in the testing protocol, these correspond to cases that are not relevant to the user. The average number of scores obtained by each variation d was assumed as the value of similarity with the reference picture q of its set.

The ground truth acquired in the comparison of each query against three sets of variations was extended to cover the comparison of each query against the overall archive through two complementary assumptions. On the one hand, we assume that the ranking obtained for variations of the same reference pictures belonging to different sets can be combined. Concretely, this means that for each reference picture, the user implicitly sets an absolute level of similarity that is maintained throughout the three subsequent sets of variations. The assumption is supported by the statistical equivalence of different sets (which are generated by a uniform random algorithm), and by the fact that different variations of the same reference picture are presented sequentially without interruption. On the other hand, we assume that any picture d_1 deriving from mutation of a reference picture q_1 has a null value of similarity with respect to any other reference picture q_2. This is to say that if d_1 would be included in a set of the reference picture q_2, then the user would rank the similarity at the lowest level. To verify the assumption, a sample of 6×9 images collecting a variation set for each reference picture was created and displayed on a page. Three pilot users were then asked to identify which variations derived from each of the six reference pictures. All the users identified a variable number of variations, ranging between four and six, with no false classifications. None of the selected images turned out to have an average rank higher than 1.2.

Based on the two assumptions, the average user-based ranking was extended to complete the array V_{qd} capturing the value of any archive picture d as a member of the retrieval set for any query q.

10.6.3 Results

Summarized in Figure 10.18 are the results of the evaluation. Reference pictures employed as queries are reported on the top, while the plots on the bottom are the curves of *Precision/Recall* obtained by resolving the query on the archive according to the joint metric of similarity based on color and weighted walkthroughs (WW), and color and centroids orientation (*CO*). Defining as *relevant* those images in the archive which are similar to the query in the users' perception, and as *retrieval set* the set of images retrieved in each retrieval session, the *Recall* is defined as the ratio between the number of relevant retrieved images and the overall number of relevant images in the archive; instead, the *Precision* is the ratio between the number of relevant retrieved images and the size of the current retrieval set.

In the plots, each point represents the values of precision and recall computed for the retrieval set which extends up to comprise the image represented by the point itself. In this way, points are added to each plot from left to right, representing retrieval sets of size varying from one image up to a maximum that depends on the specific query. In fact, the maximum size of the retrieval set for a query is determined as the number of images in the three mutation sets of the query that users recognize as similar to the query (values are 24, 24, 21, 25, 20, 23 for the six queries, respectively). In this representation, a perfect accordance of retrieval with the user ranking would result in a flat curve of precision 1 at every recall, which is possible if the retrieval set is constituted only by relevant pictures. Instead, any misclassification is highlighted by a negative slope of the curve that derives from the "anticipated" retrieval of nonrelevant pictures. For all six queries, *WW* closely fit the ideal user-based curve in the ranking of the first, and most relevant, variations. A significant divergence is observed only on the second query for the ranking of variations taking the positions between six and nine.

In all the cases tested, *WW* outperformed *CO*. In particular, *CO* evidenced a main limit in the processing of the first and the fifth queries. In particular, the long sequences with constant recall (in the case (a), the top-ranked images in the retrieval set scored a null value of recall and precision) indicate that this problem of *CO* derives from a misclassification that confuses variations of the query with those of different reference pictures. Analysis of the specific results of retrieval indicates that *CO* are not able to discriminate the first and fifth reference pictures, which are definitely different in the user perception but share an almost equal representation in terms of the centroids of color sets. Finally, note that, because all the images share a common proportion of colors, a representation based on a global histogram cannot discriminate any two images in the benchmark. As a consequence, in all the cases tested, *WW* outperformed the color histogram, which ranks, by construction, all the images in the same position.

10.6.4 A Benchmark Database of Real Images

The second stage of the evaluation was aimed at extending the experimental results to the case of images with realistic complexity. To this end, the weighted walkthroughs and the centroids orientation were experimented within the prototype system introduced in Section 10.5 and compared against a solution based on the sole use of a global color histogram. For the experiments, the system was applied to an archive of about 1000 reference paintings featured by the library of WebMuseum [89]. The test was administered to a set of 10 volunteers, all with university educations. Only two of them had experience in using systems for image retrieval by content.

Before the start of the testing phase, users were trained with two preliminary examples, in order to assure their understanding of the system. During the test, users were asked to retrieve a given set of eight target images (shown in Figure 10.19a, from T_1 to T_8), representing the aim of the search, by expressing queries by sketch (see Figure 10.19b in which the query images for a particular user are shown). To this end, users were shown each target image, and they were requested to express a query with three regions to retrieve it (see Figure 10.15). Only one trial was allowed for each target image.

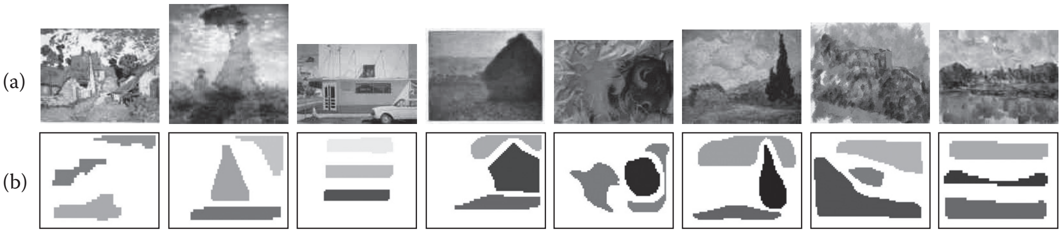


FIGURE 10.19 (See color insert.)
(a) The target images used in the test; (b) a user's query sketches for the eight images.

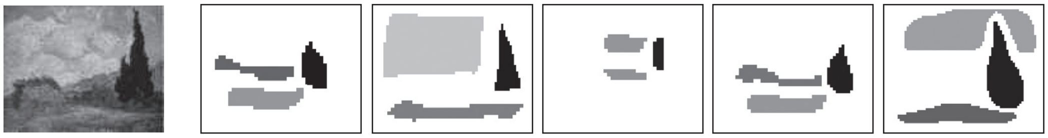
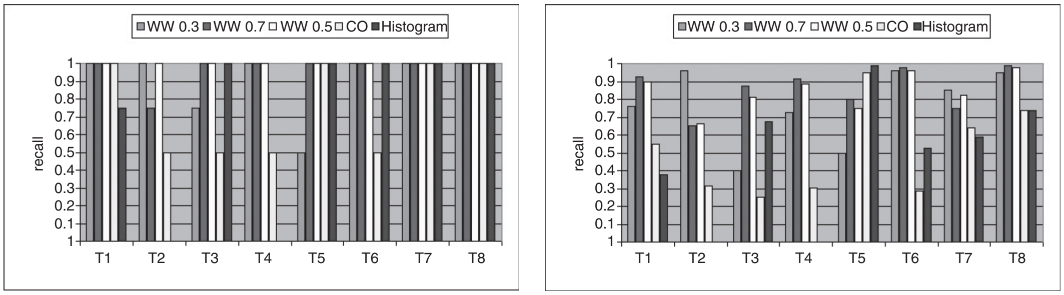


FIGURE 10.20
Different users' queries for the same target image (leftmost image).

The overall time to express queries for all eight target images was about 20 min, and this permitted the users to maintain attention. The time spent for each query, about 2 min, appeared to mainly derive from the difficulty in selecting an appropriate color capable of representing the visual appearance of the image. This is basically a limit of the interface, which is not presently engineered with respect to usability quality factors. Figure 10.20 shows some of the queries drawn from users during the search of the sixth target image. It is worth noting that different users employed noticeably different sketches to find the same target image.

For each query, the ranking of similarity on the overall set of 1000 pictures was evaluated using the joint modeling of color and spatial relationships (weighted walkthroughs and centroid orientation have been used separately to model spatial relationships), and the global color histogram. Results were summarized within two indexes of *recall* and *precision*. For each target image, the *recall* is 1 if the target image appears within the set of the first 20 retrieved images, 0 otherwise. Thus, recall expresses with a true/false condition, the presence of the target image within the retrieval set. *Precision* considers the rank scored by the target image in the retrieval set: it is 1 if the target image is ranked in the first position, and gradually decreases to 0 when the target is ranked from the first toward the 20th position (i.e., precision is assumed zero when the target is ranked out of the first 20 retrieved images). In this way, precision measures the system capability in classifying images according to the implicit ordering given by the target image. System recall and precision for each of the eight target images are derived by averaging the individual values scored for a target image on the set of users' queries.

Results are reported in Figure 10.21a and Figure 10.21b. Figure 10.21a compares values of recall for the proposed model (here indicated as *WW*), for the centroid orientation (*CO*),



(a)

(b)

FIGURE 10.21
Values of recall (a) and precision (b) are compared for the proposed model (*WW*), for centroid orientation (*CO*), and for global color histogram (*Histogram*). Results for *WW* are reported for λ = 0.5, which corresponds to an equal weight for the contribution of color and spatial distance; λ = 0.3 and λ = 0.7 correspond to a reduced or increased contribution for the color distance, respectively. It can be noticed as the global histogram definitely fails in ranking the second and fourth target images, whose recall and precision values are both null.

and for the color histogram (*Histogram*). For *WW*, results are reported for different values of the parameter λ, which weights the contribution of color and spatial distance in Equation 10.13. Though *Histogram* provides an average acceptable result, it becomes completely inappropriate in two of the eight cases (T_2 and T_4), where the recall becomes zero. Color used jointly with centroid orientation shows a recall even greater than 0.5 and performs better than *Histogram* in six of the eight cases (T_3 and T_6 are the exceptions). Differently, search based on the *WW* model provides optimal results for each of the eight tested cases. In particular, it can be observed that better results are scored for λ set equal to 0.5, while for unbalanced combinations, there are cases that both penalize a major weight for color (T_2, and this directly descends from the fact that the color histogram failed, thus evidencing the inadequacy of the sole color in obtaining the correct response for this image), and cases that penalize a major weight for spatial contribution (T_3 and T_5).

Histogram is clearly disadvantaged when the system effectiveness is measured as rank of the target image in the retrieval set, as evidenced in plots of precision of Figure 10.21b. By considering a spatial coordinate, ranking provided from the system is much closer to the user expectation, than that given by global histogram. In almost all the tested cases (T_1, T_3, T_4, and T_5), a solution that privileges the contribution of color distance scores better results than that favoring the spatial component (T_2 and T_7). In the remaining two cases (T_6 and T_8), there is no substantial difference for the three values of λ. Finally, for the target image T_5, the histogram outperforms *WW*, basically due to the low spatial characterization of this image.

10.7 Conclusions

In image search based on chromatic similarity, the effectiveness of retrieval can be improved by taking into account the spatial arrangement of colors. This can serve both to distinguish images with the same colors in different arrangements and to capture the similarity between images with different colors but similar arrangements.

In this chapter, we proposed a model of representation and comparison that attains this goal by partitioning the image in separate entities and by associating them with individual chromatic attributes and with mutual spatial relationships. Entities are identified with the sets of image pixels belonging to color clusters derived by a clustering process in the $L^*u^*v^*$ color space. In doing so, a spatial entity may be composed of multiple nonconnected segments, mirroring the human capability to merge regions with common chromatic attributes. To support this modeling approach, a suitable spatial descriptor was proposed which is able to capture the complexity of directional relationships between the image projections of color clusters.

The effectiveness of the proposed model was assessed in a two-stage experimental evaluation. In the first stage, basic chromatic arrangements were considered to evaluate the capability of the model to rank the similarity of images with equal histograms but different spatial arrangements (which cannot be distinguished using a global histogram). In the second stage, the model was experimented with to evaluate the capability to reflect perceived similarity between user sketches and images of realistic complexity. In both cases, experimental results showed the capability of the model to combine and balance account for chromatic and spatial similarity, thus improving the effectiveness of retrieval with respect to a representation based on a global histogram and a representation using centroids orientation to model spatial relationships between color clusters.

References

[1] A. Gupta and R. Jain, Visual information retrieval, *Commun. ACM*, 40, 70–79, May 1997.

[2] A. Del Bimbo, *Visual Information Retrieval*, Academic Press, San Francisco, CA, 1999.

[3] R. Veltkamp and M. Tanase, Content-Based Image Retrieval Systems: A Survey, Technical Report UU-CS-2000-34, Utrecht University, Utrecht, the Netherlands, 2002.

[4] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, Content based image retrieval at the end of the early years, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 22, 1349–1380, December 2000.

[5] T. Gevers and A. Smeulders, *Emerging Topics in Computer Visions*, ch. Content-based image retrieval: An overview. Prentice Hall, New York, 2004.

[6] T. Gevers, *Principles of Visual Information Retrieval*, ch. Color in image search engines. Springer-Verlag, Heidelberg, February 2001.

[7] R. Schettini, G. Ciocca, and S. Zuffi, *A Survey of Methods for Color Image Indexing and Retrieval in Image Databases*, ch. Color imaging science: Exploiting digital media. John Wiley & Sons, New York, 2001.

[8] C. Theoharatos, N. Laskaris, G. Economou, and S. Fotopoulos, A generic scheme for color image retrieval based on the multivariate wald-wolfowitz test, *IEEE Trans. on Knowledge and Data Eng.*, 17, 808–819, June 2005.

[9] N. Sebe and M. Lew, *Texture Features for Content-Based Retrieval*, ch. Principles of visual information Retrieval. Springer-Verlag, Heidelberg, 2001.

[10] J. Zhang and T. Tan, Brief review of invariant texture analysis methods, *Patt. Recognition*, 35, 735–747, March 2002.

[11] B. Günsel and M. Tekalp, Shape similarity matching for query-by-example, *Patt. Recognition*, 31, 931–944, July 1998.

[12] S. Loncaric, A survey of shape analysis techniques, *Patt. Recognition*, 34, 983–1001, August 1998.

[13] D. Zhang and G. Lu, Review of shape representation and description techniques, *Patt. Recognition*, 37, 1–19, January 2004.

[14] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, Object-based image retrieval using the statistical structure of images, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, Vol. 2, June 2004, pp. 490–497.

[15] S. Maybank, Detection of image structures using the fisher information and the rao metric, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 26, 1579–1589, December 2004.

[16] C. Schmid, R. Mohr, and C. Bauckhage, Evaluation of interest point detectors, *Int. J. Comput. Vision*, 37, 151–172, June 2000.

[17] D. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision*, 60, 91–110, February 2004.

[18] J.V. de Weijer and T. Gevers, Boosting saliency in color image features, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Vol. 1, June 2005, pp. 365–372.

[19] V. Gudivada and V. Raghavan, Design and evaluation of algorithms for image retrieval by spatial similarity, *ACM Trans. on Inf. Syst.*, 13, 115–144, April 1995.

[20] J. Smith and S. Chang, Visualseek: A fully automated content-based image query system, in *Proceedings of the ACM Conference on Multimedia*, Boston, February 1997, pp. 87–98.

[21] S. Berretti, A. Del Bimbo, and E. Vicario, Weighted walkthroughs between extended entities for retrieval by spatial arrangement, *IEEE Trans. on Multimedia*, 5, 52–70, March 2003.

[22] J. Amores, N. Sebe, and P. Radeva, Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Vol. 2, June 2005, pp. 769–774.

[23] M. Rodríguez and M. Jarur, A genetic algorithm for searching spatial configurations, *IEEE Trans. on Evol. Computation*, 9, 252–270, June 2005.

[24] C. Colombo, A. Del Bimbo, and P. Pala, Semantics in visual information retrieval, *IEEE Multimedia*, 6, 38–53, July–September 1999.

Au: Are these the chapter or book authors? Provide editors and page#s.

Au: Are these the chapter or book authors? Provide editors and page#s.

Au: Are these the chapter or book authors? Provide editors and page#s.

Au: Provide editors and publisher/location. Should this be Washington D.C.?

Au: Provide editors and publisher/location.

Au: Provide editors and publisher/location.

Au: Provide editors and publisher/location.

[25] B. Bradshaw, Semantic based image retrieval: A probabilistic approach, in *Proceedings of the ACM Multimedia*, Marina del Rey, October 2000, pp. 167–176.

[26] Y. Marchenco, T.-S. Chua, and I. Aristarkhova, Analysis and retrieval of paintings using artistic color concepts, in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Amsterdam, the Netherlands, July 2005, pp. 1246–1249.

Au: Provide editors and publisher/location.

[27] S. Chang and E. Jungert, Pictorial data management based upon the theory of symbolic projections, *J. Visual Languages and Computing*, 2, 195–215, June 1991.

Au: Provide editors and publisher/location.

[28] J. Smith and C.-S. Li, Decoding image semantics using composite region templates, in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, June 1998, pp. 9–13.

[29] R. Mehrotra and J. Gary, Similar-shape retrieval in shape data management, *IEEE Comput.*, 28, 57–62, September 1995.

[30] K. Siddiqi and B. Kimia, Parts of visual form: Computational aspects, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 17, 239–251, March 1995.

[31] Y. Tao and W. Grosky, Spatial color indexing: A novel approach for content-based image retrieval, in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, Vol. 1, June 1999, pp. 530–535.

[32] M. Flickner, W. Niblack, H. Sawhney, 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 Comput.*, 28, 23–32, September 1995.

Au: Provide editors and publisher/location.

[33] J. Bach, C. Fuler, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, and C. Shu, The virage image search engine: An open framework for image management, in *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases IV*, San Jose, CA, Vol. 2670, March 1996, pp. 76–87.

[34] T. Gevers and A. Smeulders, Pictoseek: Combining color and shape invariant features for image retrieval, *IEEE Trans. on Image Process.*, 9, 102–119, January 2000.

[35] C. Carson, S. Belongie, H. Greenspan, and J. Malik, Blobworld: Image segmentation using expectation maximization and its application to image querying, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 24, 1026–1038, August 2002.

[36] J. Wang, J. Li, and G. Wiederhold, Simplicity: Semantics-sensitive integrated matching for picture libraries, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 23, 947–963, September 2001.

[37] J. Li and J. Wang, Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 25, 1075–1088, September 2003.

[38] M. Swain and D. Ballard, Color indexing, *Int. J. Comput. Vision*, 7, 11–32, March 1991.

Au: Provide editors and publisher/location.

[39] Y. Rubner, C. Tomasi, and L. Guibas, A metric for distributions with applications to image databases, in *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, January 1998, pp. 59–66.

[40] A. Nagasaka and Y. Tanaka, Automatic video indexing and full video search for object appearances, in *Proceedings of the IFIP Transactions, Working Conference on Visual Database Systems II*, 1992, pp. 113–127.

Au: Provide editors and publisher/location.

[41] J. Huang, S. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, Image indexing using color correlograms, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, June 1997, pp. 762–768.

[42] J. Smith and S. Chang, Integrated spatial and feature image query, *Multimedia Syst.*, 7, 129–140, March 1999.

[43] J. Chen, T. Pappas, A. Mojsilovic, and B. Rogowitz, Adaptive perceptual color-texture image segmentation, *IEEE Trans. on Image Process.*, 14, 1524–1536, October 2005.

[44] S. Makrogiannis, G. Economou, S. Fotopoulos, and N. Bourbakis, Segmentation of color images using multiscale clustering and graph theoretic region synthesis, *IEEE Trans. on Syst., Man and Cybernetics, Part A*, 35, 224–238, March 2005.

[45] A. Del Bimbo, M. Mugnaini, P. Pala, and F. Turco, Visual querying by color perceptive regions, *Patt. Recognition*, 31, 1241–1253, September 1998.

[46] R. Haralick and L. Shapiro, Image segmentation techniques, *Comput. Vision Graphics and Image Process.*, 29, 100–132, 1985.

[47] M. Arbib and T. Uchiyama, Color image segmentation using competitive learning, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 16, 1197–1206, December 1994.

[48] D. Androutsos, K. Plataniotis, and A. Venetsanopoulos, A novel vector-based approach to color image retrieval using vector angular based distance measure, *Comput. Vision and Image Understanding*, 75, 46–58, July 1999.

[49] S. Berretti, A. Del Bimbo, and E. Vicario, Spatial arrangement of color in retrieval by visual similarity, *Patt. Recognition*, 35, 1661–1674, August 2002.

[50] G. Heidemann, Combining spatial and colour information for content based image retrieval, *Comput. Vision and Image Understanding*, 94, 234–270, April–June 2004.

[51] Multimedia Content Description Interface — part 3: Visual, Final Commitee Draft, Technical Report 15938-3, Doc. N4062, ISO/IEC, Singapore, 2001. **Au: Provide author.**

[52] J. Martinez, R. Koenen, and F. Pereira, Mpeg-7: The generic multimedia content description standard, part 1, *IEEE Trans. on Multimedia*, 9, 78–87, April/June 2002.

[53] M. Abdel-Mottaleb and S. Krishnamachari, Multimedia descriptions based on mpeg-7: Extraction and applications, *IEEE Trans. on Multimedia*, 6, 459–468, June 2004.

[54] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, Color and texture descriptors, *IEEE Trans. on Circuits and Syst. for Video Technol.*, 11, 703–715, June 2001.

[55] A. Doulamis and N. Doulamis, Generalized nonlinear relevance feedback for interactive content-based retrieval and organization, *IEEE Trans. on Circuits and Syst. for Video Technol.*, 14, 656–671, May 2004.

[56] J. Laaksonen, M. Koskela, and E. Oja, Picsom — self-organizing image retrieval with mpeg-7 content descriptors, *IEEE Trans. on Neural Networks*, 13, 841–853, July 2002.

[57] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, Relevance feedback: A power tool for interactive content-based image retrieval, *IEEE Trans. on Circuits and Syst. for Video Technol.*, 8, 644–655, September 1998.

[58] A. Kushki, P. Androutsos, K. Plataniotis, and A. Venetsanopoulos, Retrieval of images from artistic repositories using a decision fusion framework, *IEEE Trans. on Image Process.*, 13, 277–292, March 2004.

[59] A. Guttmann, R-trees: A dynamic index structure for spatial searching, in *Proceedings of the ACM International Conference on Management of Data*, Boston, MA, June 1984, pp. 47–57. **Au: Provide editors and publisher/location.**

[60] T. Sellis, N. Roussopoulos, and C. Faloustos, The r+ tre: A dynamic index for multidimensional objects, in *Proceedings of the International Conference on Very Large Databases*, Brighton, U.K., September 1987, pp. 507–518. **Au: Provide editors and publisher/location.**

[61] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger, The r* tree: An efficient and robust access method for points and rectangles, in *Proceedings of the ACM International Conference on Management of Data*, Atlantic City, NJ, May 1990, pp. 322–331. **Au: Provide editors and publisher/location.**

[62] D. White and R. Jain, Similarity indexing with the ss-tree, in *Proceedings of the IEEE International Conference on Data Engineering*, New Orleans, LA, February 1996, pp. 516–523. **Au: Provide editors and publisher/location.**

[63] N. Katayama and S. Satoh, The sr-tree: An index structure for high-dimensional nearest neighbor queries, in *Proceedings of the ACM International Conference on Management of Data*, Tucson, AZ, May 1997, pp. 369–380.

[64] M. Egenhofer and R. Franzosa, Point-set topological spatial relations, *Int. J. Geogr. Inf. Syst.*, 5, 2, 161–174, 1991.

[65] M. Egenhofer and R. Franzosa, On the equivalence of topological relations, *Int. J. Geogr. Inf. Syst.*, 9, 2, 133–152, 1995.

[66] S. Chang, Q. Shi, and C. Yan, Iconic indexing by 2-d strings, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 9, 413–427, July 1987.

[67] A. Frank, Qualitative spatial reasoning about distances and directions in geographic space, *J. Visual Languages and Computing*, 3, 343–371, September 1992.

[68] C. Freksa, Using orientation information for qualitative spatial reasoning, in *Proceedings of the International Conference on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computer Science*, Pisa, Italy, Vol. 639, 1992, pp. 162–178. **Au: Provide editors and publisher/location.**

[69] S. Berretti, A. Del Bimbo, and E. Vicario, Modeling spatial relationships between color sets, in *Proceedings of the IEEE International Workshop on Content Based Access of Image and Video Libraries*, Hilton Head, SC, June 2000, pp. 73–77. **Au: Provide editors and publisher/location.**

[70] D. Papadias and T. Sellis, The semantics of relations in 2d space using representative points: Spatial indexes, *J. Visual Languages and Computing*, 6, 53–72, March 1995.

Au: Provide publisher/location.

[71] R. Goyal and M. Egenhofer, Consistent queries over cardinal directions across different levels of detail, in *Proceedings of the International Workshop on Database and Expert Systems Applications*, A.M. Tjoa, R. Wagner, and A. Al Zobaidie, Eds., September 2000, Greenwich, U.K., pp. 876–880.

Au: Updated information available?

[72] R. Goyal and M. Egenhofer, Cardinal directions between extended spatial objects, *IEEE Trans. on Knowledge and Data Engineering* (in press).

Au: Provide editors and publisher/location.

[73] S. Cicerone and P. Di Felice, Cardinal relations between regions with a broad boundary, in *Proceedings of the ACM Symposium on Geographical Information Systems*, Washington, November 2000, pp. 15–20.

[74] S. Chang, E. Jungert, and T. Li, Representation and retrieval of symbolic pictures using generalized 2d strings, in *SPIE Proceedings of Visual Communications and Image Processing IV*, Philadelphia, Vol. 1199, November 1989, pp. 1360–1372.

[75] S. Lee and F. Hsu, Spatial reasoning and similarity retrieval of images using 2d c-strings knowledge representation, *Patt. Recognition*, 25, 305–318, March 1992.

[76] S. Lee, M. Yang, and J. Cheng, Signature file as spatial filter for iconic image database, *J. Visual Languages and Computing*, 3, 373–397, December 1992.

Au: Provide editors and publisher/location.

[77] E. Jungert, Qualitative spatial reasoning for determination of object relations using symbolic interval projections, in *Proceedings of the IEEE International Workshop on Visual Languages*, Bergen, Norway, August 1993, pp. 83–87.

Au: Provide page#s.

[78] A. Del Bimbo and E. Vicario, Specification by-example of virtual agents behavior, *IEEE Trans. on Visualization and Comput. Graphics*, 1, December 1995.

Au: Provide editors and publisher/location.

[79] D. Papadias, Spatial relation-based representation systems, in *Proceedings of the European Conference on Spatial Information Theory*, Marciana Marina, Italy, September 1993, pp. 234–247.

Au: Provide editors and publisher/location. Should this be Washington, D.C.?

[80] V. Gudivada, Spatial knowledge representation and retrieval in 3-d image databases, in *Proceedings of the International Conference on Multimedia and Computing Systems*, Washington, May 1995, pp. 90–97.

[81] K. Miyajima and A. Ralescu, Spatial organization in 2d segmented images: Representation and recognition of primitive spatial relations, *Int. J. Fuzzy Sets and Systems*, 65, 225–236, July 1994.

[82] P. Matsakis and L. Wendling, A new way to represent the relative position between areal objects, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 21, 634–643, July 1999.

[83] Y. Wang and F. Makedon, R-histogram: Qualitative representation of spatial relations for similarity-based image retrieval, in *Proceedings of the ACM Multimedia*, Berkeley, CA, November 2003, pp. 323–326.

[84] G. Dong and M. Xie, Color clustering and learning for image segmentation based on neural networks, *IEEE Trans. on Neural Networks*, 16, 925–936, July 2005.

[85] M. Eshera and K.-S. Fu, A graph distance measure for image analysis, *IEEE Trans. on Syst., Man, Cybernetics*, 14, 398–407, May/June 1984.

[86] S. Berretti, A. Del Bimbo, and E. Vicario, Efficient matching and indexing of graph models in content based retrieval, *IEEE Trans. on Patt. Anal. and Machine Intelligence*, 23, 1089–1105, October 2001.

[87] J. Ullman, An algorithm for subgraph isomorphism, *J. ACM*, 23, 31–42, January 1976.

Au: Provide editors and publisher/ location.

[88] J. Smith, Image retrieval evaluation, in *Proceedings of the IEEE Workshop of Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, June 1998, pp. 112–113.

[89] N. Pioch, WebMuseum, www.ibiblio.org/wm/, 2003.