Pattern Analysis & Applications (2001)4:83–92

Modelling Spatial Relationships between Colour Clusters

S. Berretti, A. Del Bimbo and E. Vicario

Dipartimento Sistemi e Informatica, Università di Firenze, Firenze, Italy

Abstract: Modelling of image content based on chromatic arrangement requires suitable techniques for the representation of the spatial relationship between complex sets of pixels. We propose a model of spatial relationship between extended sets, which can be computed with the same computational complexity involved in conventional representations based on centroids, but which improves effectiveness by accounting for the overall sets of pixels involved in the relationship. The effectiveness of the proposed model is compared against the orientation between centroids in a user-based evaluation.

Keywords: Image content description; Spatial relationships; Retrieval effectiveness evaluation

1. INTRODUCTION

Retrieval by visual similarity from image databases relies on models which permit us to represent image content in terms of visual features such as colour, shape or texture [1,2]. These may be referred either to the overall image, or to any subset of pixels comprising a spatial entity with high-level significance or with some low-level cohesion. When multiple entities are identified in the picture, the model may also capture relational information about their mutual spatial arrangement. In particular, integration of colour and spatial descriptors is being addressed to extend the significance of histograms with some index of spatial locality.

In a straightforward approach, this was proposed early on by Nagasaka and Tanaka [3] by partitioning images along a static grid of blocks, each associated with a separate histogram. Perceptual significance of this basic model can be improved by partitioning the image so as to fit its actual apparent patching. This is obtained automatically by clustering colour histograms around dominating components, and by identifying entities as image regions collecting *connected* pixels under common dominating colours [4,5]. Unfortunately, while few dominating colours are sufficient to partition the histogram into cohesive and significant clusters, the back-projection from the colour space to the image may split each colour into several separate regions. This commonly produces over-segmented models, which do not reflect the human capability to merge regions that are separate in space but which share common chromatic attributes.

This complexity can be circumvented by identifying entities with sets of pixels that are cohesive in the space of chromatic features, regardless of their spatial distribution in the image space [6,7]. This not only augments the perceptual robustness of representation, but also opens the way to encompass spatial relationships within more efficient matching and indexing schemes [8]. However, modelling based on this kind of entities also involves some major complexities in the representation of spatial relationships. In fact, colour clusters usually correspond to sets of pixels that are not connected, that may have small mutual distances, and may be tangled and inter-twined in complex arrangements evading any crisp classification. These complexities cannot be encompassed within common description schemes, where extended entities are replaced by the minimum embedding rectangle or by the centroid [9–12].

A representation based on the actual distribution of pixels in the set, which avoids reduction to a centroid or extension to a bounding rectangle, is proposed by Huang et al. [13] to capture the distribution of distances between pixels belonging to different colour clusters. Directional relationships between two colour clusters is represented in Smith and Chung-Sheng [14] by the frequency of different vertical displacements between dominant colours in picture samples taken along a fixed set of vertical bands. This relies on a

Received: 30 May 2000
Received in revised form: 9 February 2001
Accepted: 15 February 2001

fixed partition which does not reflect the user-perceived patching, and which thus introduces a critical trade-off between complexity and robustness of representation. In Del Bimbo and Vicario [15], this limitation is overcome by taking an integral measure of the overall quantity of pixels in two sets which are displaced along each of nine primitive directions. This results in nine values, called weighted walkthroughs, which provide a quantitative representation of the joint distribution of masses in two extended spatial entities [6]. The integral measure which underlies their definition makes weighted walkthroughs robust with respect to the complexities in the spatial distribution of entities.

The rest of the paper is organised in three sections. In Section 2, we introduce a descriptor of directional relationship between extended entities which derives from the model of weighted walkthroughs. In particular, we propose the use of a reduced but still effective set of weighted walkthroughs, we discuss their properties, and propose original algorithms for their efficient derivation. In Section 3, we evaluate the effectiveness of the descriptor, by comparing its performance against a representation based on centroid orientation and against a representation using the full set of weighted walkthroughs. Details on the process and results of the evaluation are expounded. Remarks and future work are discussed in Section 4.

2. SPATIAL RELATIONSHIPS BETWEEN PIXEL SETS

In a Cartesian reference system, a point a partitions the plane into four quadrants, *upper-left*, *upper-right*, *lower-left* and *lower-right*, that 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 pixels 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 pixels of B:

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

where $C_{\pm 1}(\cdot)$ are the characteristic functions of negative and positive real numbers, respectively, and $|B|$ denotes the area of B.

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

$$w_{i,j}(A, B) = \frac{1}{|A|\,|B|} \int_A \int_B C_i(x_b - x_a)\, C_j(y_b - y_a)\, dx_a dy_a dx_b dy_b \qquad (2)$$

In so doing, the four tuple $w(A, B)$ provides a measure of the number of pairs of pixels in A and B whose displacement falls within each of the four directional relationships. w_{11} evaluates the number of pixel pairs $a \in A$ and $b \in B$ such that b is upper right from a; in a similar manner, w_{-11} evaluates the number of pixel pairs such that b is upper left from a; w_{1-1} evaluates the number of pixel pairs such that b is lower right from a; and w_{-1-1} evaluates the number of pixel pairs such that b is lower left from a.

The four weights are adimensional positive numbers with sum equal to 1, and they are anti-symmetric (i.e. $w_{i,j}(A, B) = w_{-i,-j}(B, A)$). They are invariant with respect to shifting and zooming of the two sets A and B, and they satisfy a basic property of continuity by which small changes in the shape or arrangement of entities result in small changes of their relationships. More importantly, weights inherit from the integral operator 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:

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

If A and B are approximated by any multi-rectangular shape, the property of compositionality permits to derive the four-dimensional integrals of Eq. (2) through the linear combination of a number of closed form terms corresponding to sub-integrals taken over rectangular domains. These can be reduced to nine basic cases, represented in Figs 3(a)–(b).

2.1. Efficient Derivation

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

$$w(A, B) = w\left(\bigcup_{n=1}^{N_A} A_n, \bigcup_{m=1}^{N_B} B_m\right) = \frac{1}{|A|\,|B|} \sum_{n=1}^{N_A} |A_n| \sum_{m=1}^{N_B} |B_m| w(A_n, B_m) \qquad (3)$$

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 Fig. 3. This permits us to compute $w(A, B)$ in time $O(N * M)$.

A more elaborate strategy permits us to derive the relationship with a complexity which 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.

2.1.1. Representation of 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 directions of the Cartesian reference system. The set of cells comprising each entity is partitioned into any number of segments. Each of these segments is assumed to be connected (but not necessarily maximal with respect to the property of connection). We expound the representation of segments and the computation of their mutual relationships. Relation-

ships between the union of multiple segments is derived by direct application of the property of compositionality.

Each segment A is represented by a data structure which encompasses the following information: the number of cells of A, and the indexes i_l, j_l and i_u, j_r 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 ij in the bounding rectangle of A with a 9-tuple WW_{ij}, which encodes the number of cells of A in each of nine directions centred in the cell ij itself: WW_{ij}^{00} is equal to 1 if the cell ij is part of A, and it is equal to zero otherwise; WW_{ij}^{10} is the number of cells of A that are on the right of cell ij (i.e. the number of cells of A with indexes ik such that $k > j$); in a similar manner, WW_{ij}^{-10} is the number of cells of A that are on the left of ij, while WW_{ij}^{01} and WW_{ij}^{0-1} are the number of cells of A over and below cell ij, respectively; finally, WW_{ij}^{11} is the number of cells of A that are upper-right from ij (i.e. the cells of A with indexes hk such that $h > i$ and $k > j$); and, in a similar manner, WW_{ij}^{1-1}, WW_{ij}^{-1-1} and WW_{ij}^{-11} are the numbers of cells of A that are lower-right, lower-left and upper left from the cell i, j, respectively.

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 so doing, the nine coefficients associated to any cell ij can be derived by relying on the coefficients of the cells $i - 1$, j (lower adjacent) and i, $j - 1$ (left adjacent). The following clauses illustrate the derivation:

$$
\begin{aligned}
WW_{ij}^{00} &= 1 \text{ if the cell } ij \text{ is part of } A \\
&\quad 0 \text{ otherwise} \\
WW_{ij}^{-10} &= 0 \text{ if } j = 0 \text{ (i.e. } j \text{ is the leftmost column of } A) \\
&\quad WW_{i,j-1}^{-1,0} + WW_{ij}^{00} \text{ otherwise} \\
WW_{ij}^{10} &= \text{is derived by scanning the row } i \text{ if } j = 0 \\
&\quad WW_{i,j-1}^{1,0} - WW_{i,j}^{00} \text{ otherwise} \\
WW_{ij}^{0-1} &= 0 \text{ if } i = 0 \text{ (i.e. } i \text{ is the lowermost row of } A) \\
&\quad WW_{i-1,j}^{0,-1} + WW_{i-1,j}^{00} \text{ otherwise} \\
WW_{ij}^{01} &= \text{is derived by scanning the column } j \text{ if } i = 0 \\
&\quad WW_{i-1,j}^{01} - WW_{i,j}^{00} \text{ otherwise} \\
WW_{ij}^{-11} &= 0 \text{ if } j = 0 \\
&\quad WW_{i,j-1}^{-11} + WW_{i,j-1}^{01} \text{ otherwise} \\
WW_{ij}^{-1-1} &= 0 \text{ if } i = 0 \text{ or } j = 0 \\
&\quad WW_{i,j-1}^{-1-1} + WW_{i,j-1}^{0-1} \text{ otherwise} \\
WW_{ij}^{1-1} &= 0 \text{ if } i = 0 \\
&\quad WW_{i-1,j}^{1-1} + WW_{i-1,j}^{10} \text{ otherwise}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
WW_{ij}^{11} &= N - WW_{ij}^{00} - WW_{i,j}^{01} - WW_{i,j}^{10} \text{ if } j = 0 \text{ and } i = 0 \\
&\quad WW_{i,j-1}^{11} - WW_{i,j}^{01} \text{ if } i = 0 \text{ and } j > 0 \\
&\quad WW_{i-1,j}^{11} - WW_{i,j}^{10} \text{ if } i > 0
\end{aligned}
$$

In the overall derivation, a constant time $O(1)$ is spent for the coefficients of each cell, thus requiring a total cost $O(L_A * H_A)$, where L_A and H_A are the number 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 $O(2 * L_A * H_A)$. 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.

2.1.2. Derivation of Relationships. 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 relations of intersection between the projections of A and B on the Cartesian axes:

- If the projections of A and B have null intersection on both the axes, then the descriptor has only a non-null weight (and this weight is equal to 1) which is derived in constant time (see Fig. 3(a)).
- If the projections of A and B on the Y axis have a non-null intersection, but the projections on the X are disjoint (see, for example, Fig. 4), 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 case that the projections of A and B have non-null intersection along the X axis is managed in the same manner.

We expound here the method for the case in which A is on the left of B (see Fig. 4). In the complementary case (B on the left of A), the same algorithm serves to derive the relation $w(B, A)$, which can then be transformed into $w(A, B)$ by applying the property of anti-symmetry of weighted walkthroughs. Since all the cells of A are on the left of B, the two upper-left and lower-left weights w_{-11} (A, B) and w_{-1-1} (A, B) are equal to 0. In addition, since the sum of the four weights is equal to 1, the derivation of the upper-right weight w_{11} (A, B) is sufficient to fully determine the descriptor (as w_{1-1} $(A, B) = 1 - w_{11}$ (A, B)).

The upper-right weight w_{11} (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 Fig. 3, for any cell ij in A, the contribute to w_{11} (A, B), is equal to 1 for each cell of B having indexes hk with $h > i$ and $k > j$, and it is equal to 1/2 for each cell of B having indexes hk with $h = i$ and $k > j$. At the end of the computation, the total sum is normalised 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 right column of the

1. $UR = (A^{-1-1}_{i_l,j_A} + A^{0-1}_{i_l,j_A}) * (B^{00}_{i_l,j_B} + B^{01}_{i_l,j_B} + B^{10}_{i_l,j_B} + B^{11}_{i_l,j_B}) * 1;$
2. *for* $i = i_l : i_u$
3. $\quad UR = UR + (A^{-10}_{i_{jA}} + A^{00}_{i_{jA}}) * ((B^{00}_{i_{jB}} + B^{10}_{i_{jB}}) * 1/2 + (B^{01}_{i_{jB}} + B^{11}_{i_{jB}}) * 1);$
4. $UR = UR/(N_A * N_B);$

Fig. 1. Algorithm for the case in which A and B have null intersection along the X axis.

1. $UR = (A^{-1-1}_{i_l,j_l}) * (B^{00}_{i_l,j_l} + B^{01}_{i_l,j_l} + B^{10}_{i_l,j_l} + B^{11}_{i_l,j_l}) * 1;$
2. *for* $i = i_l : i_u$
3. $\quad UR = UR + (A^{-10}_{i_{jl}}) * ((B^{01}_{i_{jl}} + B^{11}_{i_{jl}}) * 1 + (B^{00}_{i_{jl}} + B^{10}_{i_{jl}}) * 1/2);$
4. *for* $j = j_l, ij_r$
5. $\quad UR = UR + (A^{0-1}_{i_{lj}}) * ((B^{10}_{i_l,j} + B^{11}_{i_l,j}) * 1 + (B^{00}_{i_l,j} + B^{01}_{i_l,j}) * 1/2);$
6. *for* $i = i_l i_u$
7. $\quad$ *for* $j = j_l, j_r$
8. $\quad\quad UR = UR + (A^{00}_{i,j}) * (B^{11}_{i,j} * 1 + (B^{10}_{i,j} + B^{01}_{i,j}) * 1/2 + B^{00}_{i,j} * 1/4);$
9. $UR = UR/(N_A * N_B);$

Fig. 2. Algorithm for the case in which A and B have a non-null intersection on both the axes.

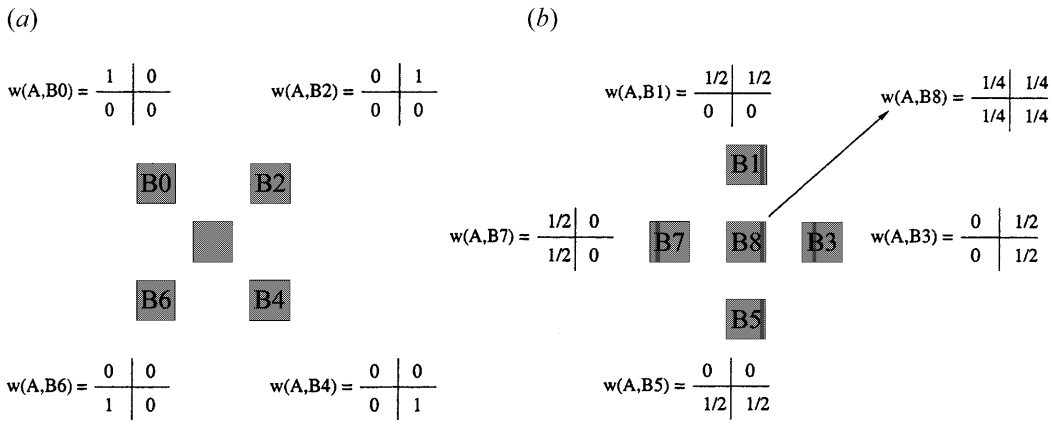

Fig. 3. The four weights for the nine basic arrangements between rectangular entities. The weights are represented as elements of a two-by-two matrix.

bounding box of A and of the left column of the bounding box of B, without covering the entire sets of cells in A and B. The algorithm is reported in Fig. 1. UR denotes the weight w_{11} (A, B) which is being computed. For the simplicity of notation, matrixes WW of segments A and B are denoted by A and B. 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 which contain cells both of A and B, respectively (see Fig. 4).

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 upper right position is equal to the term $(B^{00}_{i_l,j_B} + B^{01}_{i_l,j_B} + B^{10}_{i_l,j_B} + B^{11}_{i_l,j_B})$. According to this, statement 1, initialises UR by accounting for the contribute of all the (possibly existing) rows of A that are below row i_l. The statement in line 2, controls a loop which 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, since i_u is the topmost row of A or of B, there cannot be any other cell of A which is over row i_u, and which has any cell of B upright or aligned-right. Statement 3, in the body of the loop, adds to UR the contribute of all the cells of A belonging to row i: $(A^{-10}_{i,j_A} + A^{00}_{i,j_A})$ is the number of cells of A in the row i; each of these cells has $(B^{00}_{i,j_B} + B^{10}_{i,j_B})$ cells of B aligned on the right-hand side (contributing the weight 1/2), and $(B^{01}_{i,j_B} + B^{11}_{i,j_B})$ cells of B that are upright (each contributing the weight 1). The statement in line 4, normalises the weight.

- When projections of A and B have a non null intersection on both the axes, i.e. when the bounding boxes of A and B overlap (see Fig 5), 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.

We expound here the derivation of $w_{11}(A, B)$. Of course, any of the other three weights can be derived in a similar manner, with the same complexity. The derivation of $w_{11}(A, B)$ consists in 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 Fig. 3, each cell in the upper-right quadrant provides a contribute equal to 1, each cell in the upper column or in the right row provides a contribute equal to 1/2, and each cell coincident provides a contribute equal to 1/4.

Also in this case, matrixes WW associated with A and B permit us to accomplish the evaluation by scanning only once a limited set of cells of A and B. The algorithm is reported in Fig. 2. 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 Fig. 5).

Statement 1, initialises the weight w_{11} (A, B), denoted as UR by summing up the contribution of the $(A_{i_l j_l}^{-1-1})$ cells of A that are in the lower left quadrant of the cell i_l,j_l. 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 normalises the weight.

2.2. A Metric of Similarity

Since the four directional weights have sum equal to 1, they can be replaced without loss of information with three directional indexes, taking values within 0 and 1:

$$\begin{aligned} w_H(A, B) &= w_{1,1}(A, B) + w_{1,-1}(A, B) \\ w_V(A, B) &= w_{-1,1}(A, B) + w_{1,1}(A, B) \\ w_D(A, B) &= w_{-1,-1}(A, B) + w_{1,1}(A, B) \end{aligned} \tag{7}$$

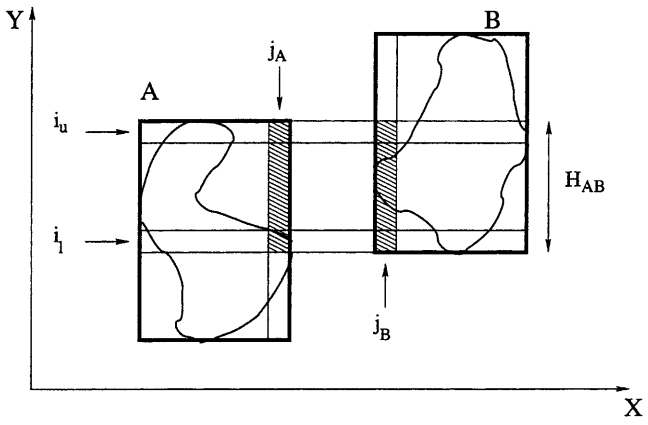

Fig. 4. Projections of bounding rectangles of A and B intersect along the Y axis. The grey patterns indicate cells that are scanned in the evaluation of the relationships.

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

Composition of differences in homologous directional indexes comprises a metric of dis-similarity $\mathcal{D}$ for the relationships between two pairs of sets $\langle A,B\rangle$ and $\langle \bar{A}, \bar{B}\rangle$ represented by weight tuples w and $\bar{w}$:

$$\mathcal{D}(w, \bar{w}) = \alpha|w_H - \bar{w}_H| + \beta|w_V - \bar{w}_V| + \gamma|w_D - \bar{w}_D| \tag{8}$$

where α, β and γ are a convex combination (i.e. they are non-negative numbers with sum equal to 1). Due to the city block structure, $\mathcal{D}$ is non-negative, auto-similar, reflexive and triangular. In addition, $\mathcal{D}$ is normal as a consequence of the bound existing on the corner weights.

3. ASSESSMENT OF EFFECTIVENESS

Perceptual significance of the metric of dis-similarity derived through the directional weights based on weighted walkthroughs was evaluated in a two-stage test, focusing first on a benchmark of basic synthetic arrangements of three colours, and then on a database of real images.

3.1. Basic Benchmark

The first stage of the evaluation was oriented to investigate the capability of weighted walkthroughs to capture the differences and similarities in basic spatial arrangements of colours, by abstracting from other relevant but independent features, such as colour distribution, size and shape of colour patches.

To this end, the evaluation was carried out on a synthetic benchmark made up of $6 \times 3 \times 9$ synthetic pictures. The archive was derived from six reference pictures, obtained by different composition of an equal number of red, yellow and blue squares within a six-by-six grid. Reference pictures (displayed on the left of the plots of Fig. 7) were created so as to contain five or six separate regions each. Preliminary pilot tests indicated that this number results in a complexity which is sufficient to prevent the user from acquiring an exact memory of the arrangement.

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 (Fig. 6 indicates the level of mutation for the nine variations in a set). 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

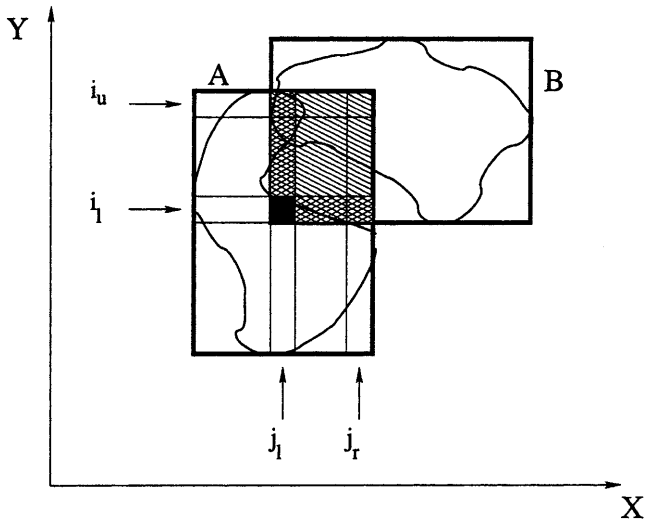

Fig. 5. Projections of bounding rectangles of *A* and *B* have a non-null intersection on both axes. During the evaluation of relationships, the cells filled with the less dense pattern are scanned once, those with a more dense pattern are scanned twice, and the black cell is scanned three times.

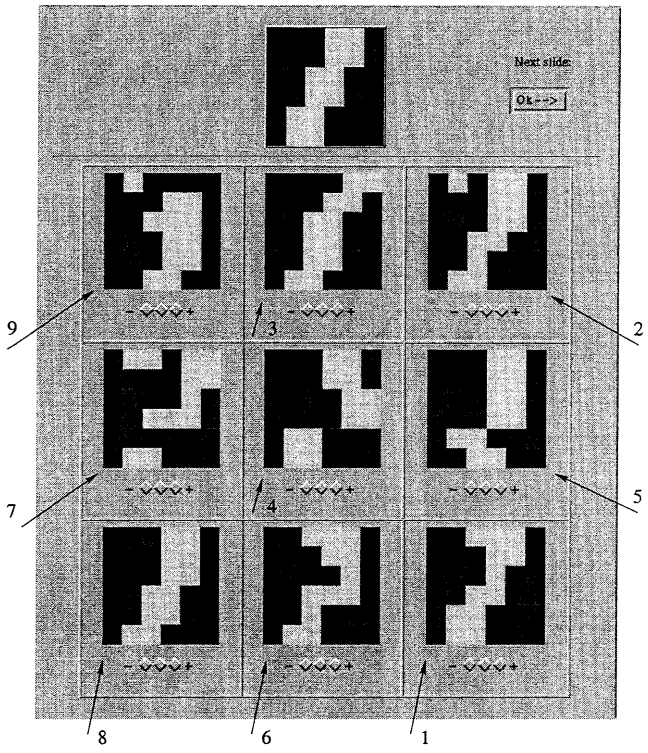

Fig. 6. A page of the user test. The numbers indicate the level of mutation.

and the multi-rectangular shape of segments, but it largely increases the fragmentation of regions. Preliminary pilot tests with variations including more than eight regions resulted in a 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. To support a competitive evaluation, queries were resolved using the metric of dis-similarity defined in Eq. (8), that comprising the complete set of weighted walkthroughs, and a conventional metric based on the difference in the orientation between the centroids of colour clusters [8]. In all the cases, the overall dis-similarity between two pictures was evaluated as the sum of dis-similarities in the three binary relationships between homologous red, yellow, and blue colour clusters.

3.2. 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 were complemented with inference.

Each user was shown a sequence of 3 × 6 html pages, each showing a reference picture and a set of nine variations. Figure 6 reports a test page, while the overall testing session is available on-line at the address http://cisc.dsi.unifi.it/forme/newtest. 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 test, users were suggested to first search for the most similar images, and then extend the rank towards low similarities, thus emphasising the relevance of high ranks.

The ground truth acquired in the comparison of each query against the three sets of its variations, was extended to cover the comparison of each query against the overall archive through two complementary assumptions.

On the one hand, we assumed 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 which 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 were presented sequentially without interruption.

On the other hand, we assumed 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 as 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 4 and 6, 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 could be extended to complete the array V_{qd}

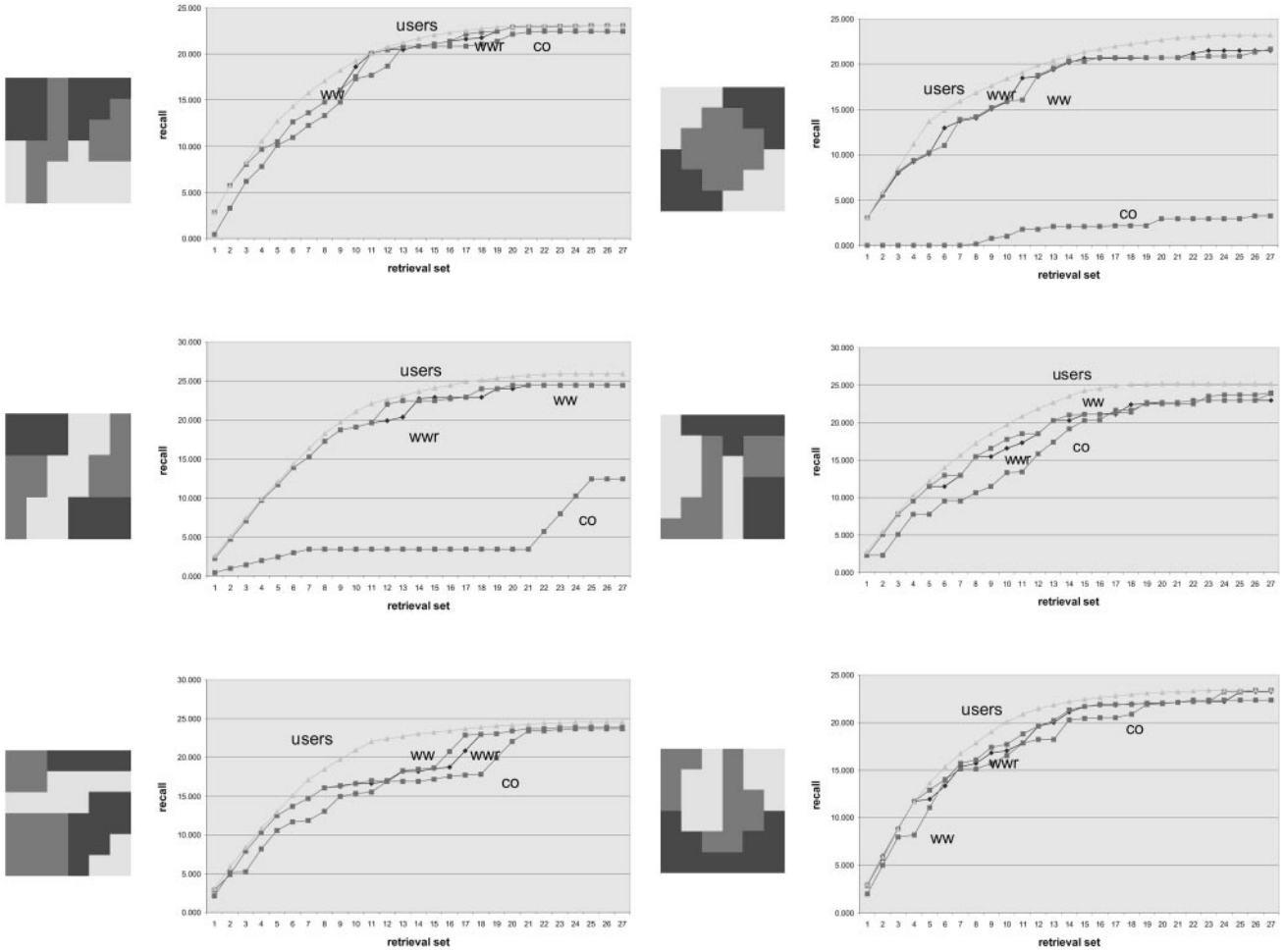

Fig. 7. The six query images and their corresponding values of recall. The recall value on the vertical axis sums up the average user-ranked value of similarity of the pictures included in the retrieval set with the dimension shown on the horizontal axis. Plotted values correspond to the ideal rank provided from the users (users), to that provided from weighted-walkthroughs (WW), reduced set of weighted-walkthroughs (WWr) and to the rank of centroid orientation (CO).

capturing the value of any archive picture d as a member of the retrieval set for any query q.

The testing session was administered to a sample of 22 volunteers, and took an average time ranging between 10 and 21 minutes, with an average of 14.6. 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 3 was attributed for each high rank received, a score of 1 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.

3.3. Results

Figure 7 summarises the results of the evaluation. Reference pictures employed as queries are reported on the left, while the plots on the right report the curves of recall and precision obtained by resolving the query on the archive according to the metrics of similarity based on weighted walkthroughs (WW), the reduced set of four directional weights (WWr) and centroid orientation (CO). The horizontal axis is the allowed dimension of the retrieval set, and the vertical axis is the sum of the values of similarity for the pictures that are included in the retrieval set. The top curve (users) resumes the ground-truth, indicating the value of recall that can be obtained with each different retrieval set when the query is managed by an ideal engine implementing the average user rank.

In this representation, a perfect accordance of retrieval with the user ranking would result in a convex slope, which is obtained when pictures are added to the retrieval set in

order of decreasing value of similarity. Any miss-classification is highlighted by a change in the convexity of the curve which derives from the 'anticipated' retrieval of a picture with lower value of similarity. Both recall and precision can be derived from the plots. In addition, the area between the users curve and that of each retrieval engine comprises a synthetic index about the effectiveness of the engine itself.

For all the six queries, WW and WWr 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 third query for the ranking of variations taking the positions between 8 and 16. In all the cases under test, WW and WWr over-perform CO. In particular, CO evidence a main limit in the processing of the second and the fourth queries. The long sequences with horizontal slope indicate that this problem of CO derives from a miss-classification, which confuses variations of the query with those of different reference pictures. Analysis of the specific results of retrieval indicate that CO is not able to discriminate the second and fourth reference pictures, which are definitely different in the user perception, but share an almost equal representation in terms of the centroids of colour clusters.

3.4. Extension to Real Images

The second stage of the evaluation was aimed at extending the comparison between weighted walkthroughs and centroid orientation to the case of images with realistic complexity. To this end, the two descriptors were experimented within a prototype system supporting retrieval by similarity based on the spatial arrangement of chromatic contents [15].

For the experiments, the system was employed on an archive developed around 200 reference paintings featured by the library of WebMuseum [17]. For each of these paintings, a set of 20 variations was derived by a random engine changing the arrangement of colour patches, so as to obtain an archive with 4200 pictures. The mutation algorithm shifts a prefixed number of chromatic patches along the horizontal, vertical, and diagonal direction, while maintaining the overall colour histogram (see Fig. 9). Also in this case, variations of the same image are derived independently, to avoid effects of ordering. The extent of mutations applied by the generation engine was set so as to ensure that the distance between mutations of the same reference image is lower than the difference between mutations of different images. According to this, a basic ground truth can be established by assuming that when any reference picture is employed as query, its variations are the 20 most relevant pictures to be retrieved.

Six reference pictures, shown in Fig. 8, were selected to be employed as queries. For each of them, the ranking of similarity on the overall set of 4200 pictures was evaluated using weighted walkthroughs and centroid orientation. Results were summarised within two indexes of *recall* and *precision*. *Recall* is evaluated as the number of mutations of the query that are ranked in the first 20 positions; *precision* is evaluated as the ratio between the number of mutations and the number of images that must be included in the retrieval set in order to retrieve all the mutations.

Results are reported in Figs 10(a)–(b). Figure 10(a) shows that *WW* score the best results, in particular recall is 1 for all tested cases. It is worth noting that centroids also provide a fair performance, due to the limited extent of mutations.

Fig. 8. The six reference images used in the experiments for retrieval and recall.

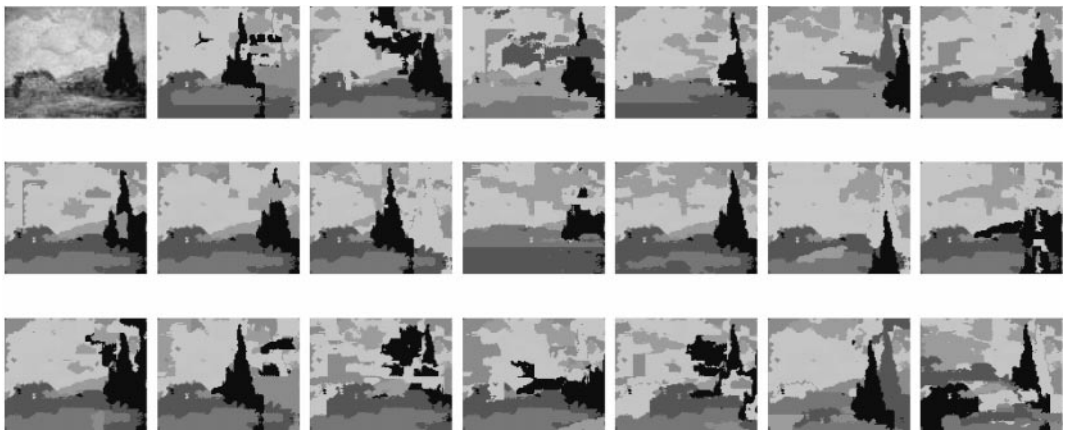

Fig. 9. The set of 20 mutations for the original painting on the upper left corner of the figure.

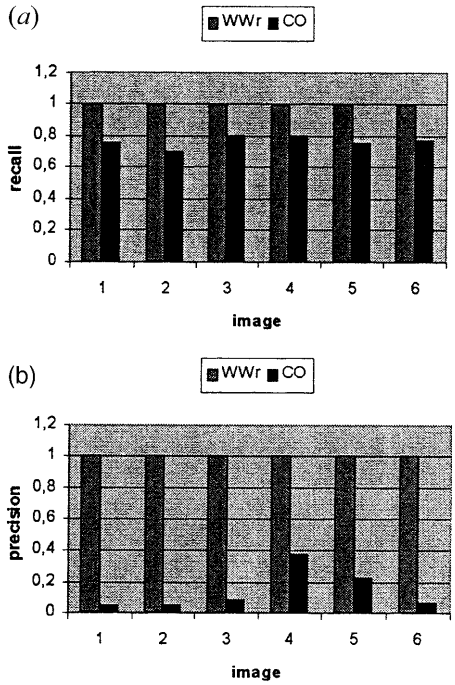

Fig. 10. Plots compare weighted walkthroughs and centroids orientation for the set of six reference images: (a) recall; (b) precision.

In fact, images sharing a similar spatial distribution of colour patches, show variations of centroids positioning that are typically small. However, using centroids orientation, arrangements occurring in images with a different visual appearance can also be assimilated as similar. This becomes evident in the plot of precision of Fig. 10(b), which shows as for centroids orientation a large retrieval set is necessary to retrieve all the mutations for a given query. Differently, weighted walkthroughs provides an optimal behaviour for all tested cases, and precision is maximum for each of the six reference images.

4. CONCLUSIONS

We proposed a model for representation of spatial relationships between extended sets of pixels developed on the weighted walkthroughs. The proposed model can be computed with the same computational complexity involved in the derivation of centroids. At the same time, the model improves the effectiveness of representation by avoiding reduction of extended sets to a single representative point, which may limit discrimination capability, especially in the management of colour clusters collecting multiple non-connected regions. A two-stages evaluation was carried out to confirm this heuristic intuition by comparing the set of directional weights against centroid orientation. In the first stage, results are acquired from a user-based test on a set of synthetic images; in the second stage measures of precision and recall are obtained from a set of real images and of their mutations.

References

1. Del Bimbo A. Visual Information Retrieval. Academic Press, 1999
2. Gupta A, Jain R. Visual information retrieval. Comm ACM 1997; 40(5):70–79
3. Nagasaka A, Tanaka Y. Automatic video indexing and full video search for object appearances. IFIP Trans Visual Database Systems II 1992; 113–127
4. Arbib M, Uchiyama T. Color image segmentation using competitive learning. IEEE Trans Pattern Analysis and Machine Intelligence 1994; 16(12):1197–1206
5. Haralick R, Shapiro L. Image segmentation techniques. Computer Vision Graphics and Image Processing 1985; 29:100–132
6. Berretti S, Del Bimbo A, Vicario E. Weighting spatial arrangement of colors in content based image retrieval. Proc IEEE Int Conf on Multimedia Comp and Sys – ICMCS'99, June 1999
7. Tao Y, Grosky W. Spatial color indexing: A novel approach for content-based image retrieval. IEEE Int Conf on Multimedia Computing and Systems, ICMCS'99, June 1999
8. Berretti S, Del Bimbo A, Vicario E. The computational aspect of retrieval by spatial arrangement. 15th International Conference on Computer Vision and Pattern Recognition (ICPR'00), September 2000
9. Chang S, Shi Q, Yan C. Iconic indexing by 2-d strings. IEEE Trans Pattern Analysis and Machine Intelligence 1987; 9(3):413–427
10. Chang S, Jungert E. Pictorial data management based upon the theory of symbolic projections. J Visual Languages and Computing 1991; 2(2):195–215
11. Smith J, Chang S. Visualseek: a fully automated content-based image query system. ACM Multimedia '96, November 1996
12. Gudivada V, Raghavan V. Design and evaluation of algorithms for image retrieval by spatial similarity. ACM Trans Information Systems 1995; 13(2)
13. Huang J, Kumar S, Mitra M, Zhu W-J, Zabih R. Image indexing using color correlograms. IEEE Conference on Computer Vision and Pattern Recognition June 1997; 762–768
14. Smith J, Chung-Sheng L. Decoding image semantics using composite region templates. IEEE CVPR98 Workshop on Content-based Access to Image and Video Libraries, June 1998
15. Del Bimbo A, Vicario E. Using weighted spatial relationships in retrieval by visual contents. IEEE Workshop of Content-Based Access of Image and Video Databases, June 1998
16. Smith J. Image retrieval evaluation. IEEE Workshop of Content-Based Access of Image and Video Databases, June 1998
17. WebMuseum. http://www.oir.ucf.edu.

Stefano Berretti received his doctoral degree in electronics engineering in 1997 from the Università degli Studi di Firenze, Italy. He is currently a PhD candidate in Information Technology and Communications at the same university. His current research interest is mainly focused on content modelling and retrieval for image and video databases.

Alberto Del Bimbo is Full Professor and Director of the Department of Sistemi e Informatica at the Università degli Studi di Firenze, Italy. He is also the Director of the Master in Multimedia at the same university. His scientific interests and activity have addressed the subject of image technology and multimedia, with particular reference to object recognition and image sequence analysis, content-based retrieval for image and video databases, visual languages and advanced man-machine interaction. Professor Del Bimbo is the author of over 150 publications, has appeared in the most distinguished international journals and conference proceedings, and is the author of the monography *Visual Information Retrieval* edited by Morgan Kaufman in 1999. He has also been the Guest Editor of several special issues of international journals and the chairman of several conferences

in the field of image processing, image databases and multimedia. He is an IAPR fellow and presently a Member of the Steering Committee of IEEE ICME, Int. Conference on Multimedia and Expo and of the VISUAL conference series. From 1996 to 2000 he was the President of the Italian Chapter of IAPR, the International Association for Pattern Recognition. Since 1999 he has been a Member of the IEEE Publications Board. He presently serves as Associate Editor of IEEE Transactions on Multimedia, IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, the Journal of Visual Languages and Computing, and Multimedia Tools and Applications Journal.

Enrico Vicario received the doctoral degree in electronics engineering and the PhD in information technology and communications from the Università di Firenze, in 1990 and 1994, respectively. Since 1998, he has been Associate Professor of Information Engineering at the Universities of Ancona and Florence. The scientific research of Enrico Vicario developed on both software engineering and visual information technology. In this latter field, his current interest is mainly focused on content modelling and retrieval for image and video databases. Enrico Vicario is an Associate Editor of *IEEE Multimedia*.

Correspondence and offprint requests to: A. Del Bimbo, Dipartimento Sistemi e Informatica, Università di Firenze, via S. Marta 3, 50139 Firenze, Italy. E-mail: delbimbo@dsi.unifi.it