Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 939–949
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Retrieval of objects in video by similarity based on graph matching

F. Chevalier *, J.-P. Domenger, J. Benois-Pineau, M. Delest

Laboratoire Bordelais de Recherche en Informatique, Université des Sciences et Technologies de Bordeaux, 33405 Talence Cedex, France

Received 11 April 2005; received in revised form 22 November 2006
Available online 31 December 2006

Communicated by S. Dickinson

Abstract

In this paper, we tackle the problem of matching of objects in video in the context of the rough indexing paradigm. The approach developed is based on matching of region adjacency graphs (RAG) of pre-segmented objects. In the context of the rough indexing paradigm, the video data are of very low resolution and segmentation is consequently inaccurate. Hence the RAGs vary with the time. The contribution of this paper is a graph matching method for such RAGs based on an improvement of relaxation labelling techniques. In this method, adjustments of similarity between regions according to neighborhood consistency compensate for the inaccuracy of segmentation. The approach demonstrates promising performance on real sequences when compared to another region-based technique.

Keywords: Video object matching; Rough indexing paradigm; Relaxation; Inexact graph matching; CBIR

1. Introduction

This paper addresses the problem of object retrieval in video, and more precisely, matching of a moving object extracted from prototype video frame with objects extracted from other frames in a video stream. Typical applications of our method are the retrieval of objects in video-shot collections or grouping of shots containing the same protagonist into video scenes.

In video, the shape, the size and the structure of objects change mainly due to camera motion, object motion and occlusion phenomena. Thus, the structure of the same object at different times in a video may present significant differences.

Furthermore, our work is placed in the context of the rough indexing paradigm (Erol and Kossentini, 2000, 2001; Seales et al., 1998). This is a new trend of fast and approximate multimedia indexing. It implies that the multimedia (video in our case) data is available at low resolution, e.g. it comes from partially decoded MPEG compressed streams. The downsampling produces a smoothing of colorimetric and geometrical data. It brings supplementary noise into results of segmentation methods applied to such data. Consequently, our algorithm aims to retrieve objects that are very similar and reject objects that differ strongly, but a large uncertainty persists for objects of intermediate similarity. We note that such kinds of application are still very poorly addressed in literature, apart from works of Erol and Kossentini (2000, 2001). In their case, object shapes are already encoded as a part of MPEG-4 standard compressed stream. Here, the object matching is done on the basis of shape descriptors (Erol and Kossentini, 2000). DC coefficients of color blocks are used to form the color histograms of compressed data in order to compare shaped MPEG-4 video objects (Erol and Kossentini, 2001). These methods are thus based on global features of objects and do not take into account their structure.

In this paper, we propose comparison of articulated objects resulting from region-based and motion-based segmentation of video frames at a low resolution. An overview

* Corresponding author. Tel.: +33 5 40 00 69 00; fax: +33 5 40 00 66 69.
E-mail addresses: chevalie@labri.fr (F. Chevalier), domenger@labri.fr (J.-P. Domenger), benois-p@labri.fr (J. Benois-Pineau), maylis@labri.fr (M. Delest).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.12.009

of the method is displayed in Fig. 1. A segmented object is represented by a RAG (*Step* 1 in Fig. 1). This classical model of representation allows us to encode the region adjacency relations (Conte et al., 2004; Gomila and Meyer, 2003). Therefore, we can express the matching of segmented objects in terms of graph matching. As the partitions of the same object may strongly differ with time in video due to its motion, occlusions and segmentation noise, the corresponding RAGs may be strongly different as well. Consequently, an exact graph matching is not possible (Conte et al., 2004). Techniques for inexact or error-tolerant graph matching are frequently used in content-based image retrieval (CBIR) and are more adequate for our purpose (Huet and Hancock, 1999; Llados et al., 2001; Shapiro and Haralick, 1981; Wilson and Hancock, 1997; Wilson, 1996). These methods aim to make a correspondence between the nodes of the graphs without imposing finding a graph isomorphism. Techniques based on Ullman's algorithm start with a single vertex to vertex mapping and then gradually extend this matching while it fulfills matching constraints (Ullman et al., 1976). When the matching does not respect the constraints, the process backtracks. Although this kind of approach is particularly efficient for trees, it has high complexity on graphs (Dinitz et al., 1999; Ullman et al., 1976). Another error-tolerant graph matching approach for RAG matching would consider a similarity measure between the regions of objects based on region features. It would build a complete bipartite graph with the sets of nodes consisting of all nodes of the two RAGs to be matched. The edges of the bipartite graph are weighted by the similarity measures between pairs of nodes. The maximum cardinality, maximum weighted coupling computed on this graph induces a set of correspondences between regions of the two objects. In this case, even if we maximize the matching between pairs of regions, the topology of the objects is not taken into account. Thus, this approach may produce some matching mistakes because of the loss of the objects' structure information.

Recently, many-to-many graph matching has been studied in application to object recognition problem (Demirci et al., 2006; Dickinson et al., 2005; Keselman and Dickinson, 2005). In this framework, the restrictive assumption of one-to-one node correspondence is overcome to compensate segmentation errors, object articulation, scale difference and within-class deformation. However, in the worst case, any subset of nodes in one graph can match any subset of nodes in the other and the space of possible many-to-many correspondences between the graphs is exponentially growing. Demirci et al. have proposed transforming the graphs into points in a low-dimensional geometric space using low-distortion graph embedding techniques (Demirci et al., 2006). Each point in the embedding space corresponds to a node in the original graphs. The distance in the embedding space reflects the shortest-path distance in the original graphs in order to keep topological relations. Assuming that attributes of a node can be mapped to a vector of masses, the many-to-many vector correspondences are mapped back into many-to-many correspondences between graph nodes. In this way, the many-to-many problem is tractable (Demirci et al., 2006). This method is very promising and can be applied with success to our problem. Nevertheless, we think that classical one-to-one matching methods such as relaxation labelling technique are still of much interest, even in case of segmentation errors and natural variation of regions due to the articulated motion of objects formed by these regions. To overcome these problems, we propose a simplification of RAGs to compensate such segmentation noise. This simplification is depicted in Fig. 1 by *Step* 2 block.

Stochastic relaxation techniques introduced a long time ago (Hummel and Zucker, 1983; Kittler et al., 1985; Rosenfeld et al., 1976) for pattern recognition have recently received new interest in several multimedia applications (Gomila and Meyer, 2003; Wing Hing Kwan et al., 2001). In the problem of object matching in video, natural objects are often articulated and even if region characteristics vary with time, the structure of a region neighborhood would remain stable. This is why relaxation labelling techniques are justified. Based on a similarity measure computed between pairs of regions, relaxation processes aim to introduce evaluation of local neighborhood likenesses to adjust the similarity measure between pairs of regions. In this

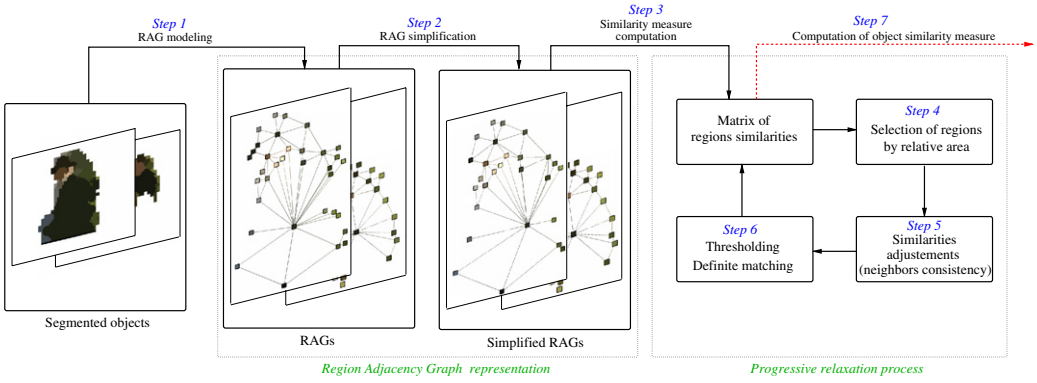


Fig. 1. The overall scheme of object matching method.

way, the regions of an object are recognizable even if small local motions of the object or segmentation errors have deformed them.

The graph matching method we present is inspired by relaxation methods (Hummel and Zucker, 1983; Rosenfeld et al., 1976). A similarity measure between the nodes of RAGs is computed (*Step 3* in Fig. 1). Then, this similarity measure is iteratively updated according to the contribution of local neighborhoods (*Step* 5). This measure may increase or decrease depending on the neighborhood resemblance. We propose to start the relaxation process with the most important regions in terms of surface (number of pixels). Hence it will be ensured that segmentation noise often expressed by small regions would not affect matching of significant regions in objects. Thus matching process starts with larger regions and introduces small ones at each step as is depicted in *Step* 4 in Fig. 1. Contrary to the approach by relaxation techniques proposed by Gomila for tracking of objects (Gomila and Meyer, 2003), in our case of matching of objects independently segmented in video frames, the variation of RAGs can be significant. Thus, starting matching of RAGs from largest regions will help us to make the matching process more robust.

The paper is organized as follows. In Section 2, we briefly introduce segmentation of objects in the rough indexing paradigm and describe how RAGs are built. In Section 3, we introduce similarity measures between two regions. Section 4 describes the preliminary step for the relaxation process we propose, that is RAG simplification (*Step* 2 in Fig. 1). The RAG matching algorithm is described in Section 5. Results on natural video are presented in Section 6 and a conclusion is given in Section 7.

2. Segmentation and RAG-modeling of objects from "rough" video

In this paper, we only consider DC-spatial resolution of video frames (Manjunath et al., 2002). The DC-images are composed of color pixels which represent the mean values of 8×8 squared blocks in original video frames. In this way, the colorimetric and geometrical information is strongly smoothed.

The segmentation process used in this work is based on a region growing algorithm performed with a modified watershed (Manerba et al., 2005). It produces a partition into 4-adjacent regions that represent a segmented object $\mathcal{O}$. We recall that two regions are called 4-adjacent if, when modeling pixels as square boxes, they share a border segment of at least one pixel and not just a pixel vertex. Each region is homogeneous according to a colorimetric homogeneity criterion which expresses the difference of color vectors of pixels in a region and the mean color vector of a region compared to a region adaptive threshold [Manerba et al., 2005]. In a classical way, we associate a partition $\mathcal{O} = \{r_1, \ldots, r_n\}$ to a RAG denoted by $\mathcal{G}_\mathcal{O}$ (see *Step* 1 in Fig. 1). Each region $r_i \in \mathcal{O}$ is considered as a vertex of $\mathcal{G}_\mathcal{O}$ and there exists an edge $e = (r_i, r_j)$ between two vertices if the corresponding regions r_i and r_j are 4-adjacent. Then, we note by $\mathcal{N}_\mathcal{O}(r_i)$ the neighborhood of r_i (i.e. the set of regions r_j that are 4-adjacent to r_i).

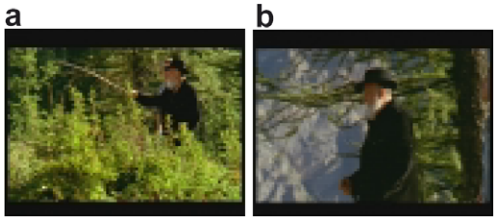


Fig. 2. Original two frames from a video stream at low resolution.

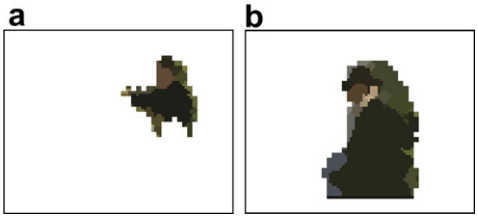


Fig. 3. Extracted objects.

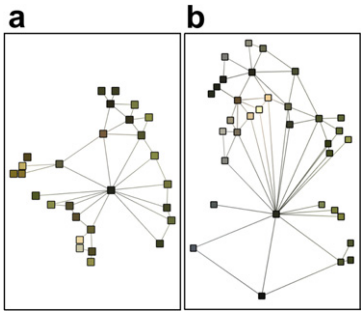


Fig. 4. Region adjacency graphs associated with the objects in Fig. 3: (a) 28 vertices; (b) 34 vertices.

In Fig. 2, two video frames at different times are shown. The same object (an old man) appears in both frames. The results of the object segmentation are displayed in Fig. 3. One can see that many differences exist due to scale deformation, local motions (e.g. the man's arm), partial occlusion and additional background pixels. The corresponding graphs are displayed[1] in Fig. 4. Here, each node of a RAG is represented by a squared box centered at a region's center of gravity. The boxes are filled in with the mean colors of corresponding regions in image plane. The edges depict regions' adjacency.

3. Features extraction and region similarity measure

In order to define similarity between graphs, we use a set of features to describe an object. Most approaches consider color as the major querying feature. Shape and texture are other usual features in the CBIR framework.

[1] The RAGs are drawn with a graph visualization tool named Tulip (Auber, 2003).

We assume that region-based object retrieval implies that local properties of regions are taken into consideration. This means that objects are represented by a set of regions characterized by local features and topological relations instead of a summarized global features vector.

The chosen descriptors have to be adapted according to the data resolution. In this work, data has low colorimetric and geometrical resolution. Therefore, the use of sophisticated features (e.g. normalized MPEG-7 color and shape descriptors (Manjunath et al., 2002)) is inappropriate for our application. Texture information is not exploitable because the downsampling attenuates the texture information. Then, our choice is to consider basic color and shape features to characterize regions. These features are the mean color vector of a region in the RGB color system and geometrical features defined by the oriented bounding box of a region in the image plane. Hence, a comparison of regions becomes a comparison of their feature vectors (see *Step* 3 in Fig. 1).

In this paper, we assume that if two regions strongly differ on one of the features, they are not similar. We call this property the *absorbing property*. This means that for the regions to be similar, they have to be close for both color and shape features accordingly to a defined similarity measure. We assume that if two regions are very much similar in shape but are very distant in color, they are less similar than two regions that have close values in both shape and color. Obviously, these values should not be too "small".

3.1. Color similarity

Color is the most frequently used feature for querying and retrieving multimedia data. Color histograms are common tools in image and video retrieval (Erol and Kossentini, 2001; Flickner et al., 1995; Qi and Han, 2005; Saykol et al., 2005). Distance in different color spaces is also a classical similarity measure in this application domain (Gomila and Meyer, 2003; Li et al., 2000; Wang et al., 2001). In this work the segmentation process aggregates DC-blocks according to an homogeneity criterion. Indeed, the color histogram of a region is not really relevant as the size of regions is small. Consequently, we consider the simplest feature, such as the mean color vector of a region in the RGB space as the colorimetric feature. We define the color similarity measure between two regions r and r' as $\rho_c(r,r')$:

$$\rho_c(r,r') = \left(1 - \frac{|\overline{R}(r) - \overline{R}(r')|}{255}\right) \cdot \left(1 - \frac{|\overline{G}(r) - \overline{G}(r')|}{255}\right) \cdot \left(1 - \frac{|\overline{B}(r) - \overline{B}(r')|}{255}\right). \quad (3.1)$$

Here, $(\overline{R}(r), \overline{G}(r), \overline{B}(r))^T$ is the mean color vector of a region r in the RGB color system.

The colorimetric criterion verifies the absorbing property. This means that a strong difference for one of the color components induces a strong decrease of the global color similarity measure. A value close to 1 indicates a high similarity.

3.2. Shape similarity

Shape characterization of the regions is another feature that is often considered in CBIR. Shape representations can be divided into two categories: a region representation and a boundary representation (Huang and Rui, 1997; Zhang and Lu, 2004). Region features are used to characterize the inside of a region. Moment invariants, bounding box features (e.g. compactness, elongatedness, eccentricity) are examples of descriptors of this type. A second approach considers boundaries of regions. Fourier descriptors and MPEG-7 contour shape descriptors are examples of descriptors belonging to this category (Manjunath et al., 2002). In DC-images, boundaries are particularly smoothed because of the block-based summarization of full resolution images. Thus, boundary descriptors are less appropriate than region descriptors to describe the regions in our context.

Our shape descriptor has been chosen to be invariant with respect to the usual transformations such as rotation, translation and scaling. Moreover the shape of many articulated objects changes between two different frames because of local motion of object parts (e.g. the arm of the old man in Fig. 3). Thus, some regions may be deformed in a way different from the others. Therefore, the shape similarity measure must be tolerant to geometrical deformations of regions. In the case of a scale deformation, the whole object is scaled with the same zoom factor. Then, we also consider the relative area of regions according to the whole surface of the object. We use the region's oriented bounding box (OBB) properties to characterize a region shape. Let r be a region, we consider its eccentricity $e(r)$ (ratio between the surface of the region and the surface of its corresponding OBB), its elongatedness $l(r)$ (ratio between the major and the minor axes of the OBB) and its relative area $a(r)$ (ratio between the surface of the region and the whole surface of the object). The shape similarity measure between two regions r and r' is as follows:

$$\rho_s(r,r') = \frac{1}{2} \cdot \left(\frac{\min(e(r),e(r'))}{\max(e(r),e(r'))} + \frac{\min(l(r),l(r'))}{\max(l(r),l(r'))}\right) \cdot \frac{\min(a(r),a(r'))}{\max(a(r),a(r'))}. \quad (3.2)$$

This shape similarity measure benefits from the absorbing property: regions have to be close from both local (OBB) and global (relative area) points of view to be considered as similar.

3.3. Global region similarity measure

Based on the shape and color similarity, we introduce a global similarity measure of regions. Let r and r' be two

regions, we will denote the similarity between two regions by $\rho(r,r')$ defined as follows:

$$\rho(r,r') = \rho_c(r,r') \cdot \rho_s(r,r'). \quad (3.3)$$

The value $\rho(r,r')$ is in $[0,1]$. The closer to 1 is the similarity measure, the more similar are the regions. When two regions differ in one feature, they are not similar. The absorbing property implied by multiplicative scheme (3.3) is then justified.

4. Simplification of a region adjacency graph

The downsampling introduced by DC-images with regard to full resolution frames may produce superfluous regions in a partition. Indeed, a pixel in a DC-image corresponds to an 8×8 block of pixels in the full resolution frame. The value of a DC-pixel is a mean value of this 8×8 block. Thus, if a 8×8 block in a full resolution frame contains portions of regions very different in color, then its corresponding pixel in DC-resolution frame will be a mix of these colors. Therefore it will strongly differ from surrounding DC-pixels corresponding to homogeneous 8×8 blocks inside regions. In this case small regions may appear around the boundary of the object. In order to avoid these drawbacks and improve the recognition of objects, we have introduced a preliminary step to simplify the RAGs (*Step* 2 in Fig. 1). We merge the regions resulting from over-segmentation.

When merging, we distinguish three cases: (i) very small regions that have to be systematically merged with their neighbors, (ii) large regions that are preserved and (iii) regions of medium size resulting from over-segmentation which are merged only if there exists a region in their neighborhood with a close color. Here, the color of a region r' is considered close to color of a region r if the similarity measure $\rho_c(r,r')$ (see Eq. (3.1)) is above a given threshold $\delta \in [0,1]$.

More formally, let $\varepsilon_1 < \varepsilon_2$ be relative area thresholds in $[0,1]$. Let us now consider each region r according to its relative area $a(r)$. Let the region r' be the closest colorimetric neighbor of r (i.e. $r' = \max_{r_i \in \mathcal{N}_\mathcal{O}(r)} \rho_c(r_i, r)$). Then the following merging rules are applied:

(1) if $a(r) < \varepsilon_1$ then the region r is removed by merging r with r';
(2) if $a(r) \geqslant \varepsilon_2$, then r is a relevant region and nothing is done;
(3) if $\varepsilon_1 \leqslant a(r) < \varepsilon_2$, if $\rho_c(r,r') > \delta$ we consider r as an artifact of over-segmentation and r is removed by merging r with r'.

In Fig. 5, we present the simplified graphs obtained by the reduction of the graphs displayed in Fig. 4. The red[2] arrows show merged regions.

[2] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.

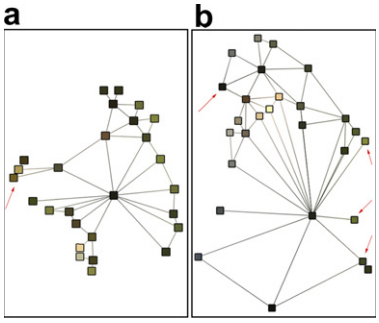


Fig. 5. Reduction of the graphs of the Fig. 4. Merged regions are depicted by red arrows: (a) 27 vertices; (b) 27 vertices.

5. Object matching with RAGs

The matching process is based on an iterative relaxation process that progressively refines the similarity between the regions by taking into account the similarity of their neighborhood. In such an approach, the evaluation of the similarity between the regions is obtained by combining initial similarity (based on features) and successive adjustments (reinforcement or penalization) according to the local context information (i.e. regions of the neighborhood).

5.1. Overview of the algorithm

The largest regions form the most significant parts of the objects, their similarity measure is consequently the most relevant in terms of importance and accuracy. It is necessary to match them as well as possible without being disturbed by small neighboring regions. On the contrary, small regions may result from segmentation errors. In this case, they penalize the similarity of the large regions located in their neighborhood. Therefore, we propose an *ordered matching* of regions in RAGs. Starting with two simplified RAGs (Section 4 and *Step* 2 in Fig. 1), we first compute the initial similarity measures of all regions pairwise (*Step* 3 in Fig. 1). Then, the algorithm proceeds iteratively.

Let μ_i be a threshold inversely proportional to i, the iteration number. At each iteration i, we only consider regions of objects that have a relative area above the threshold μ_i (*Step* 4 in Fig. 1). This means that, from an object $\mathcal{O}$, we compute the reduced object $\mathcal{O}_i$ such that $r \in \mathcal{O}_i$ iff $a(r) > \mu_i$. In the same manner, we reduce $\mathcal{O}'$ into $\mathcal{O}'_i$. Thus, starting with the largest regions, we introduce smaller regions at each iteration in the matching process.

At ith iteration, for each pair of regions (r,r') of reduced objects $\mathcal{O}_i$ and $\mathcal{O}'_i$, we compute the adjusted similarity $\gamma_i(r,r')$ between r and r'. This adjusted similarity depends on the initial similarity measure and on the similarity of their neighborhood at $i-1$th iteration (*Step* 5 in Fig. 1). We fix two similarity thresholds $\theta_{\min} < \theta_{\max}$. If the value $\gamma_i(r,r')$ is lower than $\theta_{\min}$, we assume that regions will never match and then we assign $\gamma_i(r,r')$ to zero. In an opposite way, if the value $\gamma_i(r,r')$ is stronger than $\theta_{\max}$ and presents

no ambiguity with other regions we decide that the regions have definitively matched and we assign $\gamma_i(r,r')$ to one (*Step* 6 in Fig. 1).

In order to accelerate the convergence of the relaxation algorithm, we limit the number of iterations to t loops. Our experience is that increasing t above 10 does not improve the matching result. The algorithm is the following, the corresponding steps in Fig. 1 are indicated in italic.

Here, a pair of regions is "marked" when its similarity measure has been assigned to one or zero. The neighborhood compensation function will be described in Section 5.2.

5.2. Adjustment by neighborhood contribution

The neighborhood contribution (see *Step* 5 in Fig. 1) denoted by φ_i is based on the similarity between the restricted neighborhood $\mathcal{N}_{\mathcal{O}_i}(r)$ of the region r and the restricted neighborhood $\mathcal{N}_{\mathcal{O}'_i}(r')$ of the region r'. The similarity measure $\gamma_i(r,r')$ is updated depending on φ_i in the following way:

$$\gamma_i(r,r') = \rho(r,r') + \varphi_i(\mathcal{N}_{\mathcal{O}_i}(r), \mathcal{N}_{\mathcal{O}'_i}(r')). \tag{5.1}$$

The adjustment φ_i is used to increase or decrease the similarity between two regions according to the confidence of their neighborhoods $\mathcal{N}_{\mathcal{O}_i}(r)$ and $\mathcal{N}_{\mathcal{O}'_i}(r')$. We will first introduce the neighborhood similarity measure denoted by $\kappa_i(r,r')$. Its computation is based on the best coupling that can be found between the neighbors of r and the neighbors of r'. We build the complete bipartite graph composed of regions of the neighborhood of r and r' respectively and weighed by the similarity measures $\gamma_{i-1}(r,r')$ between regions. Then, we compute the maximum cardinality, maximum flow $\mathcal{F}$ mapping on the complete bipartite graph (Hopcroft and Karp, 1973; Weber and Mlivoncic, 2003). This algorithm has been used by Shokoufandeh et al. in the framework of object recognition (Shokoufandeh et al., 1999). In their work, objects are defined at a different level of abstraction. The maximum cardinality, minimum weight algorithm is iteratively computed on the graphs at each level of abstraction.

Fig. 7 shows a bipartite neighborhood graph that corresponds to $\mathcal{N}_{\mathcal{O}_i}(r) = \{r_0, r_1\}$ and $\mathcal{N}_{\mathcal{O}'_i}(r') = \{r'_0, r'_1, r'_2\}$ and the similarity evaluation $\gamma_{i-1}(r_k, r'_l)$ Fig. 8.

In the definition between the neighborhood of r and r' we propose to take into account their relative cardinality and their best matching by average max flow $\overline{\mathcal{F}}$. Hence, the similarity measure will take into account the similarity of the structure of neighborhoods and the similarity of the regions in these neighborhoods in the descriptor space. Thus, neighborhood similarity $\kappa_i(r,r')$ is given by the following formula:

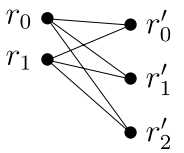
$$\kappa_i(r,r') = \frac{\min(|\mathcal{N}_{\mathcal{O}_i}(r)|, |\mathcal{N}_{\mathcal{O}'_i}(r')|)}{\max(|\mathcal{N}_{\mathcal{O}_i}(r)|, |\mathcal{N}_{\mathcal{O}'_i}(r')|)} \cdot \overline{\mathcal{F}}. \tag{5.2}$$

It decreases if the structure of the neighborhoods strongly differs or if the regions are different in the feature space.

The neighborhood similarity $\kappa_i(r,r')$ will now be used to define the neighborhood compensation function $\varphi_i(\mathcal{N}_{\mathcal{O}_i}(r), \mathcal{N}_{\mathcal{O}'_i}(r'))$ in Eq. (5.1). The role of $\varphi_i(\mathcal{N}_{\mathcal{O}_i}(r), \mathcal{N}_{\mathcal{O}'_i}(r'))$ (see Eq. (5.3)) is to adjust the similarity between two regions according to their neighborhood similarity. In order to ensure a significant influence of the neighborhood similarity or dissimilarity on the adjustment process, we consider three cases of adjustment according to the value of $\kappa_i(r,r')$: (i) we increase the similarity between the regions if their neighborhoods strongly correspond, (ii) we decrease it if the neighborhoods are very different and (iii) the neighborhoods are not similar or dissimilar enough to take a decision and we do not modify the similarity between regions. Let α_1 and $\alpha_2 \in [0,1]$ be the thresholds to differen-

```
globalAlgorithm (O, O', t) Step 7
{
  initRegionSimilarities(O, O') Step 3
  for i from 1 to t
     (O_i, O'_i) ← selectRegions(O, O', i) Step 4
     for each pair (r, r')_{r∈O_i, r'∈O'_i}
        if (r, r') is marked
          γ_i(r, r') ← γ_{i-1}(r, r')
        else
          γ_i(r, r') ← neighborhoodCompensation(r, r', O_i, O'_i) Step 5
          γ_i(r, r') ← thresholding(γ_i(r, r')) Step 6
        endif
     endfor
     matching(O_i, O'_i, O, O') Step 6
  endfor
}
```

Fig. 6. Global matching algorithm.



γ_{i-1}	r'_0	r'_1	r'_2
r_0	0.5	0.2	0.1
r_1	0.6	0.4	0.2

Fig. 7. The bipartite graph and the table of similarities.

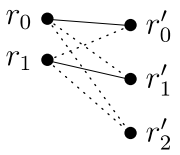


Fig. 8. Maximal weighted coupling.

tiate these cases. The value of $\varphi_i(\mathcal{N}_{\mathcal{O}_i}(r), \mathcal{N}_{\mathcal{O}'_i}(r'))$ is defined as

$$\varphi_i(\mathcal{N}_{\mathcal{O}_i}(r), \mathcal{N}_{\mathcal{O}'_i}(r')) = \begin{cases} \dfrac{1}{2\alpha_1}\kappa_i(r,r') - \dfrac{1}{2} & \text{if} \quad \kappa_i(r,r') \leqslant \alpha_1, \\ \dfrac{1}{2(1-\alpha_2)}\kappa_i(r,r') - \dfrac{\alpha_2}{2(1-\alpha_2)} & \text{if} \quad \alpha_1 < \kappa_i(r,r') \leqslant \alpha_2, \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

In practice, we split the interval of possible values of κ equally. This means that $\alpha_1 = \frac{1}{3}$ and $\alpha_2 = \frac{2}{3}$.

5.3. Matching of regions

If, during the matching process, due to relaxation (5.3) or not, the similarity of a pair of regions has became very strong or very low, we definitively match them. Thus, the global matching process will be speeded up as we exclude these regions from matching of RAGs.

Matching of RAGs is speeded up by including two steps. First, we fix a null similarity value between regions that will probably never match (thresholding in *Step* 6 in the global algorithm). Secondly, we consider that two regions match each other if the similarity value is high enough and there is no ambiguity with other regions (i.e. there are no other potential candidates for matching with one or other region). This step corresponds to the *Step* 6 in the overview of the method (Fig. 1).

Let θ be the matching threshold, and let ε be the ambiguity threshold. The conditions for a definite matching denoted by $r \approx r'$ between two regions r and r' are defined as follows:

$$r \approx r' \Longleftrightarrow \begin{cases} \gamma(r,r') > \theta, \\ \gamma(r,r') - \gamma(r,r'_j) > \varepsilon \quad \forall r'_j \neq r' \in \mathcal{O}', \\ \gamma(r,r') - \gamma(r_i,r') > \varepsilon \quad \forall r_i \neq r \in \mathcal{O}. \end{cases} \tag{5.4}$$

The first condition in (5.4) stands for a high similarity value. The second and the third conditions ensure that there are no any other potential candidates for matching, that is "no ambiguity". In practice, we fix $\varepsilon = 0.08$ and $\theta = 0.8$.

The limitation of matching by these (5.4) aims to avoid abundance of bad correspondences. The conditions (5.4) seem strong for the regions to be in correspondence, but neighborhood consistency adjustments along iterations help pairs of corresponding regions to fulfill these constraints after they have been in an ambiguous position at previous iterations.

If a pair of regions (r,r') verifies $r \approx r'$ (see property (5.4)), a definite match can be done. Then, for all $j \geqslant i$ we have

$$\begin{aligned} &\gamma_j(r,r') = 1, \\ &\forall r_k \in \mathcal{O}, \quad r_k \neq r, \quad \gamma_j(r_k,r') = 0, \\ &\forall r'_l \in \mathcal{O}', \quad r'_l \neq r', \quad \gamma_j(r,r'_l) = 0. \end{aligned}$$

The matching process described above is computed by the "*matching*" procedure of the global algorithm in Fig. 6 and it corresponds to the *Step* 6 of the overview of the method displayed in Fig. 1.

5.4. Object similarity

The similarity measure we introduce between the objects corresponds to the average of the relative areas that have been matched. Let R and R' be the sets of regions of $\mathcal{O}$ and $\mathcal{O}'$ respectively that have been matched. At the end of the matching process (*Step* 7 in Fig. 1), the similarity $\sigma(\mathcal{O},\mathcal{O}')$ between $\mathcal{O}$ and $\mathcal{O}'$ is computed as follows:

$$\sigma(\mathcal{O},\mathcal{O}') = \frac{1}{2}\left(\sum_{r\in R} a(r) + \sum_{r'\in R'} a(r')\right). \tag{5.5}$$

Here, $a(r)$ is the relative surface of a region r with regard to the object surface as introduced in Section 3.2.

The similarity measure σ evaluates the mean proportion of object areas that have been matched. This means that we first compute the whole relative area of matched regions for each set R and R'. The object similarity measure corresponds to the mean of these two values.

5.5. Study of the complexity of the global matching algorithm

The complexity of the matching algorithm can be assessed as follows. At each iteration of the outer loop, we consider all pairs (r,r') from the reduced objects. Let $\overline{V}$ and $\overline{V'}$ be the average cardinals of $\mathcal{O}_i$ and $\mathcal{O}'_i$ over all iterations $i = 1,\ldots,t$. The inner loop is run through $\overline{V}\overline{V'}$ times. For each run of the inner loop, we consider the neighborhoods of r and r'. Let $\overline{v}$ and $\overline{v'}$ be the average number of neighbors of r and r'. The computation of the neighborhood compensation uses the maximal weighted matching algorithm applied on the complete bipartite graph of neighborhoods. This algorithm runs in $(\overline{v}+\overline{v'})^{\frac{5}{2}}$ operations (Hopcroft and Karp, 1973). Thus, the average complexity

of the global algorithm is $\Theta(t\overline{V}\overline{V'}(\overline{v}+\overline{v'})^{\frac{5}{2}})$. We recall that t is a fixed number of iterations of the outer loop.

The best-case complexity will thus be in a constant time $(\Omega(t2^{\frac{5}{2}}))$, and the worst-case complexity will be $O\left(tnm[(n-1)+(m-1)]^{\frac{5}{2}}\right)$, with n and m the cardinals of node sets of $\mathscr{G}_{\mathscr{O}}$ and $\mathscr{G}'_{\mathscr{O}}$ respectively. We note that the worst-case is not realistic as it corresponds to a complete RAG case. The latter is not possible for planar segmentation maps with the number of regions higher than 3.

6. Experimental results and discussion

We have tested our method for retrieval of objects in sequences at DC-resolution taken from CERIMES ©MPEG2-compressed documentaries. The segmented objects are extracted from DC-frames of size 76×92 pixels and at the temporal resolution of two frames per second.

The sequences are taken from CERIMES ©documentary videos *Aquaculture en méditerranée*, *De l'arbre à l'ouvrage*, *Le chancre* and *Hiragasy*. Fig. 9 shows an overview of the shots used for experiments. The whole video database contains about 5000 frames from which objects are extracted. For the experiments, 100 objects corresponding to people have systematically been chosen randomly from the video objects database.

We have evaluated the performance of our method in the context of object retrieval by query by example. Retrieval systems often present query by example results in terms of k best matches (Flickner et al., 1995; Pentland et al., 1996; Qi and Han, 2005; Wang et al., 2001). A retrieved object is considered a correct match if it represents the same object as the query. Two examples of object retrieval are shown in Fig. 10. The scores under frames correspond to the object similarity measure σ. The example (a) illustrates the ability of our method to retrieve the same object under different conditions: the similarity measures are good even if the same old man appears in two different shots. In the example (b), the two first retrievals are relevant with a similarity measure over 0.7 while other retrievals do not correspond to the query with a similarity measure "close to chance" ($\simeq$0.5).

In order to prove the interest of considering neighborhoods for matching process as in our method, we have compared our relaxation process with the IRM method used in the SIMPLIcity system (Li et al., 2000; Wang et al., 2001). Both algorithms have been implemented using only our own region similarity measure (3.3) with a color feature as a vector $(\overline{R},\overline{G},\overline{B})$ (3.1) and shape features $e(r)$, $l(r)$ and $a(r)$ (3.2). Because of the roughness of data, the texture is smoothed and the color moments of higher order than mean are not representative. Thus, IRM which is based on these sophisticated features becomes less efficient. In this way, IRM is strongly penalized as shown by the results in Fig. 11.

The precision figures for different values of the number of best matches k for both methods are plotted in Fig. 12. Precision is computed as being the ratio between the number of correct matches and k. Whereas the slope is the same for both methods, our method is more precise by about 20%. The score of IRM is less than ours because it is penalized by the incompleteness of the region features set. Our method seems to better compensate the poorness of the information contained in regions' features by considering the topology of objects with regions' neighborhoods.

In Fig. 13, the precision corresponds to the ratio of correct matches versus total number of objects whose score σ

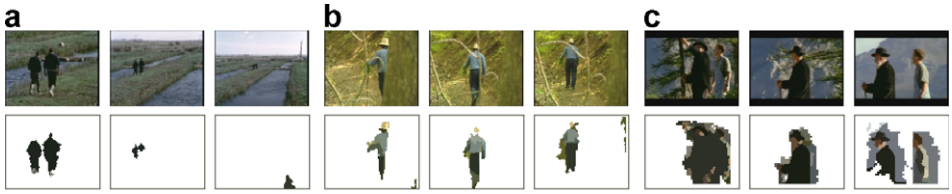


Fig. 9. Sample of shots taken from the video database: (a) example shot 1; (b) example shot 2; (c) example shot 3.

Fig. 10. Our method: global object similarities with the query (left frame with a red border) for the five best retrievals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. IRM: global object similarities with the query (left frame with a red border) for the five best retrievals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

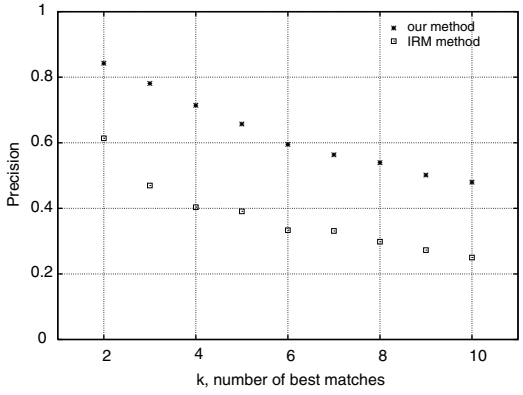


Fig. 12. Object retrieval precision for different values of k.

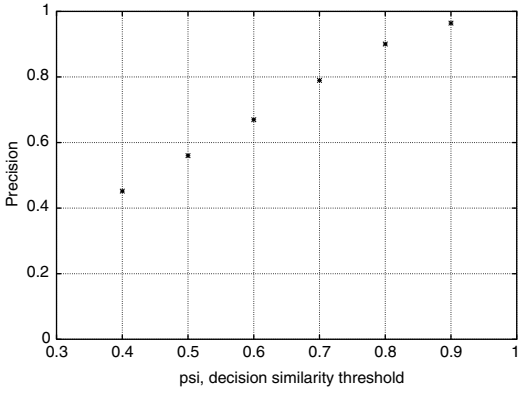


Fig. 13. Object retrieval precision for different values of ψ.

is over a fixed threshold ψ. The precision for IRM was not computed as this method requires richer descriptor space, otherwise the similarity measure (5.5) is close to 1 even for a bad matching. The curve we obtain is strongly increasing as the threshold raises. For a ψ value fixed around 0.7, the precision is quite good with a corresponding recall of 0.6.

We have tested the robustness of the method to image alterations. Figs. 14 and 15 summarize the results. The graphs in Fig. 14 show the precision on the five best retrievals as we increase the significance of image alterations.

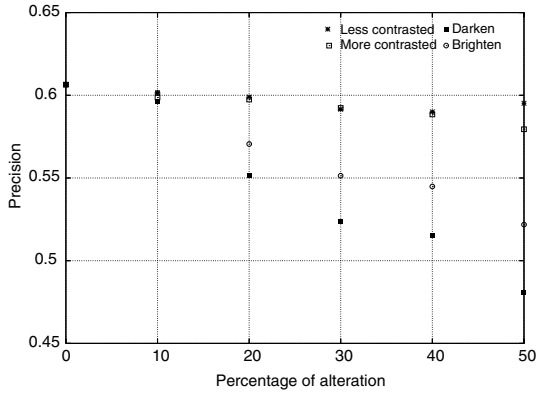


Fig. 14. The robustness of the method to contrast and intensity variations. Precision on five best matches.

The method is extremely robust to contrast variation as shown in Fig. 14. The method is also stable to intensity variation. The Fig. 15 shows some query examples to images alterations such as intensity variation, contrast variation, random noise, zoom and rotation.

7. Conclusion

Thus in this paper we have presented a new approach to the problem of object matching recognition in video in the context of the rough indexing paradigm inspired by relaxation labelling approaches for graph matching. The method relies on a similarity-based region matching. It represents an improvement in relaxation techniques for video objects at a low resolution modeled by RAGs. Information is scarce due to the down-sampling. Thus, classical methods of content-based recognition that frequently use sophisticated region features such as texture are penalized due to the inaccuracy of such data.

In this paper, we have proposed a new region similarity measure adapted to the data, and a graph simplification technique to eliminate regions resulting from segmentation noise.

Another contribution was to propose a progressive matching of regions-nodes of graphs accordingly to their relative surface. This new approach combined with a thres-

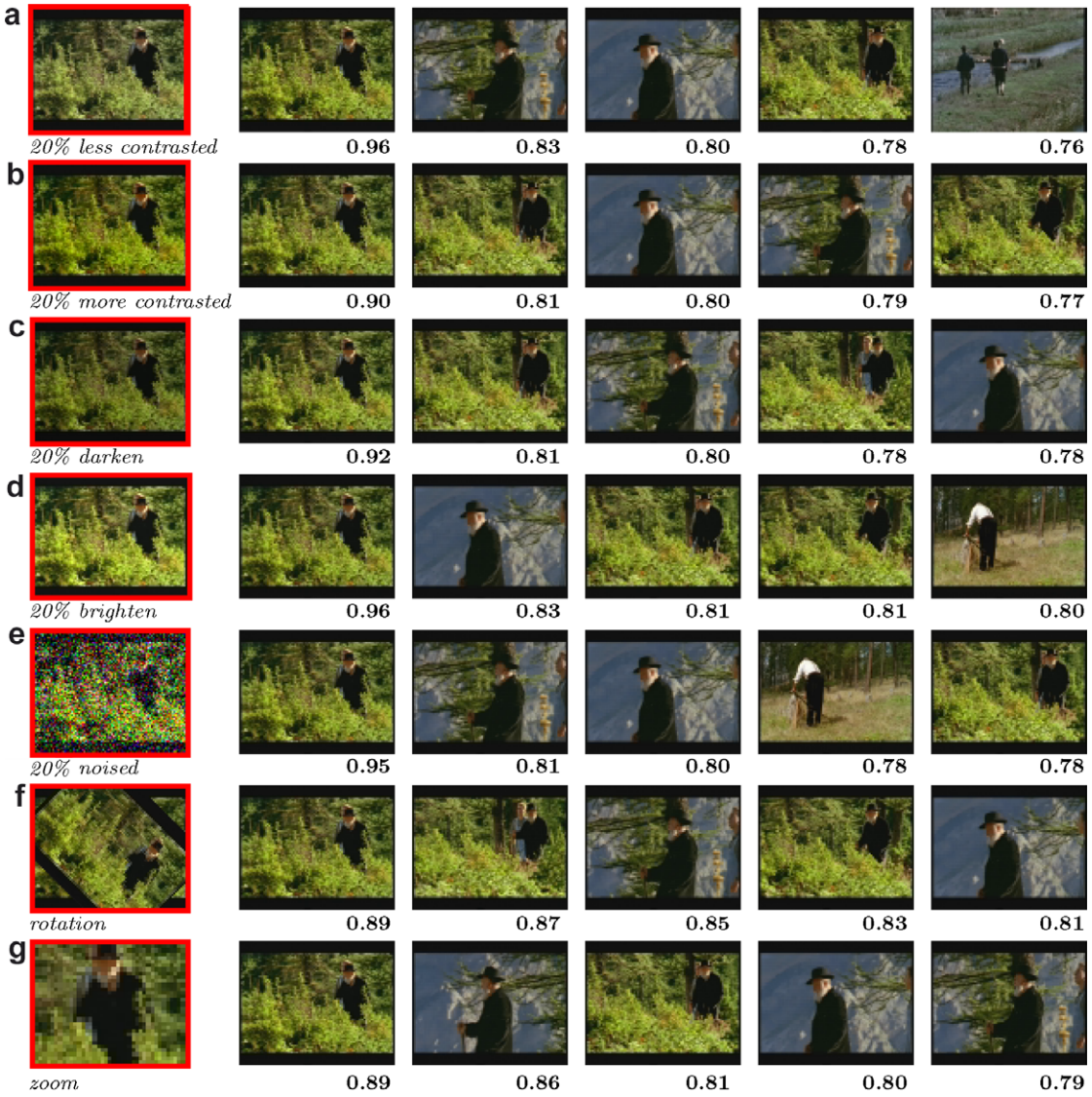


Fig. 15. The robustness of the method to image alteration. Best five matches are shown.

holding process permits the algorithm to better match the seed regions in the context of the neighborhood adjustment. We thus privileged the most relevant regions in order to better match less important areas.

As an immediate application of our method we see the retrieval of shots in a video database. In this scenario, we search for shots containing the same object as the query. A video database contains a selection of k representative frames per shot with associated pre-computed RAGs of objects. By submitting a query object to the database, we select all objects such that the similarity measure with the query is above a threshold ψ. Thus, the corresponding shots can be selected as a response to query. Another application is a semantic inventory of the shots in a video into video chapters or scenes. Here, the shots will be grouped into a scene if they content the same object as a "query shot".

In the case of a large database, a filtering step based on global features of objects such as global histogram can be combined with our matching method in order to reduce the number of objects to test (Qi and Han, 2005).

As a final conclusion, we can say that this approach offers good results and a nice challenge for further improvements.

References

Auber, D., 2003. Tulip – a huge graph visualization framework. In: Graph Drawing Software. Springer-Verlag.

Conte, D., Foggia, P., Sansone, C., Vento, M., 2004. Thirty years of graph matching in pattern recognition. Internat. J. Pattern Recognit. Artif. Intell. 18 (3), 265–298.

Demirci, M.F., Shokoufandeh, A., Keselman, Y., Bretzner, L., Dickinson, S., 2006. Object recognition as many-to-many feature matching. Internat. J. Comput. Vision 69 (2), 203–222.

Dickinson, S., Shokoufandeh, A., Keselman, Y., Macrini, D., 2005. Object categorization and the need for many-to-many matching. In: Proc. 27th DAGM – The Annual meeting of the German Association for Pattern Recognition, Vienna, Austria, August.

Dinitz, Y., Itai, A., Rodeg, M., 1999. On an algorithm of Zemlyachenko for subtree isomorphism. Inform. Process. Lett. 70 (3), 141–146.

Erol, B., Kossentini, F., 2000. Retrieval of video objects by compressed domain shape features. In: The 7th IEEE International Conference on Electronics, Circuits and Systems, 2. Jounieh, Lebanon, pp. 667–670.

Erol, B., Kossentini, F., 2001. Color content matching of mpeg-4 video objects. PCM '01: Proc. of the Second IEEE Pacific Rim Conf. on Multimedia. Springer-Verlag, pp. 891–896.

Flickner, M., Sawhney, H., Niblack, W., et al., 1995. Query by image and video content: The qbic system. IEEE Computer 28, 23–32.

Gomila, C., Meyer, F., 2003. Graph-based object tracking. In: Internat. Conf. on Image Processing September 14–17.

Hopcroft, J., Karp, R., 1973. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. SIAM J. Comput. 2, 225–231.

Huang, T., Rui, Y., 1997. Image retrieval: Past, present, and future. Proc. Internat. Symposium on Multimedia Information Processing, 1–23.

Huet, B., Hancock, E.R., 1999. Inexact graph matching. In: IEEE CVPR99 Workshop on Content-based Access of Image and Video Libraries (CBAIVL-99), Fort Collins, Colorado USA, June 22.

Hummel, R.A., Zucker, S.W., 1983. On the foundations of relaxation labeling processes. IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-5 (3), 267–287.

Keselman, Y., Dickinson, S., 2005. Generic model abstraction from examples. IEEE Trans. on Pattern Analysis and Machine Intelligence 27 (7), 1141–1156.

Kittler, J., Christmas, W.J., Petrou, M., 1985. Probabilistic relaxation for matching problems in computer vision. IEEE Trans. on Pattern Analysis and Machine Intelligence 7 (5), 617–623.

Li, J., Wang, J.Z., Wiederhold, G., 2000. IRM: integrated region matching for image retrieval. Proc. Eighth ACM Multimedia Conf., 147–156.

Llados, J., Marti, E., Villanueva, J.J., 2001. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. IEEE Trans. Pattern Anal. Machine Intell. 23 (10), 1137–1143.

Manerba, F., Benois-Pineau, J., Leonardi, R., 2005. Real-time rough extraction of foreground objects in MPEG1,2 compressed video. In: Proc. Workshop on Image Analysis For Multimedia Interactive Services (WIAMIS), April, Montreux, Switzerland.

Manjunath, B., Salembier, P., Sikora, T. (Eds.), 2002. Introduction to MPEG-7. Wiley.

Pentland, A., Picard, R., Sclaroff, S., 1996. Photobook: Content-based manipulation of image databases. Internat. J. Comput. Vision 18 (3), 233–254.

Qi, X., Han, Y., 2005. A novel fusion approach to content-based image retrieval. Pattern Recognition 38, 2449–2465.

Rosenfeld, A., Hummel, R., Zucker, S.W., 1976. Scene labeling by relaxation operations. IEEE Trans. Systems Man and Cybernetics SMC-6 (6), 420–433.

Saykol, E., Güdükbay, U., Ulusoy, Ö., 2005. A histogram-based approach for object-based query-by-shape-and-color in image and video databases. Image Vision Comput. 23 (13), 1170–1180.

Seales, W., Yuan, C., Hu, W., Cutts, M., 1998. Object recognition in compressed imagery. Image Vision Comput. 16 (5), 337–352.

Shapiro, L., Haralick, R., 1981. Structural descriptions and inexact matching. IEEE Trans. Pattern Anal. Machine Intell. 3 (5), 504–519.

Shokoufandeh, A., Marsic, I., Dickinson, S., 1999. View-based object recognition using saliency maps. Image Vision Comput. 17 (5–6), 445–460.

Ullman, J.R., Sridhar, V., Li, X., 1976. An algorithm for subgraph isomorphism. J. ACM 23 (1), 31–42.

Wang, J.Z., Li, J., Wiederhold, G., 2001. SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries. IEEE Trans. Pattern Anal. Machine Intell. 23 (9), 947–963.

Weber, R., Mlivoncic, M., 2003. Efficient region-based image retrieval. Image Vision Comput. 21 (3), 285–294.

Wilson, R.C., 1996. Inexact Graph Matching Using Symbolic Constraints. PhD thesis, The University of York, November.

Wilson, R., Hancock, E., 1997. Structural matching by discrete relaxation. IEEE Trans. Pattern Anal. Machine Intell. 19, 634–648.

Wing Hing Kwan, P., Kameyama, K., Toraichi, K., 2001. Trademark retrieval by relaxation matching on fluency function approximated image contours. In: IEEE Pacific Rim Conf. on Comm., Comp. and Sig. Pro., pp. 255–258.

Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognition 1, 1–19.