Unsupervised watershed-driven region-based image retrieval

I. Pratikakis, I. Vanhamel, H. Sahli, B. Gatos and S.J. Perantonis

Abstract: A novel unsupervised strategy for content-based image retrieval is presented. It is based on a meaningful segmentation procedure that can provide proper distributions for matching via the earth mover's distance as a similarity metric. The segmentation procedure is based on a hierarchical watershed-driven algorithm that extracts meaningful regions automatically. In this framework, the proposed robust feature extraction and the many-to-many region matching along with the novel region weighting for enhancing feature discrimination play a major role. Experimental results demonstrate the performance of the proposed strategy.

1 Introduction

Increasing amounts of imagery because of advances in computer technologies and the advent of world wide web have made apparent the need for effective and efficient imagery indexing and retrieval based not only on the metadata associated with it (e.g. captions and annotations) but also directly on the visual content. During the evolution period of content-based image retrieval (CBIR) research, the major bottleneck has been the gap between low-level features and high-level semantic concepts. Therefore the obvious effort toward improving a CBIR system is to focus on methodologies that will enable a reduction or even, in the best case, bridging of the aforementioned gap. Image segmentation plays a key role toward the semantic description of an image, as it provides the delineation of the objects that are present in an image. Although contemporary algorithms cannot provide a perfect segmentation, some can produce a rich set of meaningful regions upon which robust discriminant regional features can be computed.

This paper presents a strategy for CBIR. It is based on a meaningful segmentation procedure that can provide proper distributions for matching via the earth mover's distance (EMD) as a similarity metric. The segmentation procedure relies on a hierarchical watershed-driven algorithm that extracts meaningful regions automatically. In this framework, the proposed robust feature extraction along with a novel region weighting that enhances feature discrimination play a major role. The complete process for querying and retrieval does not require any supervision by the user. The user's only interaction is the selection of an example image as query. Experimental results demonstrate the performance of the proposed strategy.

2 Related work

The fundamental aspects that the existing region-based image retrieval systems take into consideration are the following: (i) the segmentation scheme; (ii) the selected features for region representation; (iii) the region matching method and (iv) the user supervision.

The NeTra system [1] is presented where retrieval is based on segmented image regions. The segmentation scheme requires user supervision for parameter tuning and segmentation corrections. Furthermore, a one-to-one region matching is proposed after region selection by the user. In the same spirit, the Blobworld system is proposed by Carson *et al.* [2], in which a user is required to select important regions and features. As an extension to Blobworld, Greenspan *et al.* [3] compute blobs by using Gaussian mixture modelling and use EMD [4] to compute both the dissimilarity of the images and the flow-matrix of the blobs between the images.

Fuh *et al.* [5] use the idea of combining colour segmentation with relationship trees and a corresponding matching method. They use information concerning the hierarchical relationship of the regions along with the region features for a robust retrieval. An integrated matching algorithm is proposed by Wang *et al.* [6] which is based on region similarities with respect to a combination of colour, shape and texture information. The proposed method enables one-to-many region matching. Hsieh and Grimson [7] propose a framework that supports a representation for a visual concept using regions of multiple images. They support one-to-many regions matching in two stages. First, a similarity comparison occurs followed by a region voting that leads to a final region matching. Mezaris *et al.* [8] propose an approach that employs a fully unsupervised segmentation algorithm and associate low-level descriptors with appropriate qualitative intermediate-level descriptors, which form a simple vocabulary termed object ontology. Following that, a relevance feedback mechanism is invoked to rank the remaining, potentially relevant image

IEE Proceedings online no. 20050061

doi:10.1049/ip-vis:20050061

Paper first received 27th February and in revised form 19th July 2005

I. Pratikakis, B. Gatos and S.J. Perantonis are with the Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research 'Demokritos', Athens 153 10, Greece

I. Vanhamel and H. Sahli are with the Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels 1050, Belgium

E-mail: ipratika@iit.demokritos.gr

regions and produce the final query results. Finally, Jing *et al.* [9] propose an image retrieval framework that integrates efficient region-based representation and effective on-line learning capability. This approach is based on user's relevance feedback that makes user supervision an obligatory requirement.

In this paper, unlike the above approaches, we propose a strategy that does not require any supervision from the user apart from selecting an example image to be used as a query and permit a many-to-many region matching improving the robustness of the system. It is a region-based approach that takes advantage of the robustness of each subsequent module. More specifically, it is based on a watershed-driven hierarchical segmentation module that produces meaningful regions, and a feature extraction module that expresses meaningful distributions for matching along with a robust similarity metric that is fed with a novel weighting factor.

3 Image representation

3.1 Automatic multiscale watershed segmentation

The proposed watershed-driven hierarchical segmentation scheme is based on a modified version of an image segmentation approach for vector-valued images presented previously by Vanhamel *et al.* [10] and Vanhamel *et al.* [11]. It consists of three basic modules that are preceded by a step that determines whether texture features should be taken into account in the segmentation process (Fig. 1). The first module (salient measure module) is dedicated to a scale-space analysis based on multiscale watershed segmentation and nonlinear diffusion filtering. This module creates a weighted region adjacency graph (RAG), in which the weights incorporate the notion of scale. Using the obtained multiscale RAG, the second module (hierarchical level selection module) extracts a set of partitionings that have different levels of abstraction, denoted as hierarchical levels. The last module (segmentation evaluation module) identifies the most suitable hierarchical level for further processing, which in this work corresponds to the level containing all significant image features. The remaining of the section is structured as follows. First, we discuss the selection of the feature-space required for the segmentation process. Next, we explain the salient measure module, in which we comment on the employed nonlinear diffusion and the creation of the multiscale RAG. Finally, we discuss the concept of hierarchical level selection, and the definition and selection of the most suitable level.

3.1.1 Feature-space selection for image segmentation: To accommodate for texture, the segmentation scheme can be applied on a colour–texture feature space [10]. Spectral decomposition is a common way to describe texture in image processing. The texture content is usually represented as a vector-valued image, in which each decomposition band describes the energy at a given frequency and orientation. The spectral decomposition using Gabor filtering has often been justified by the fact that it provides a good approximation of the natural processes in the primary visual cortex. A Gabor function is a harmonic wave modulated by a Gaussian. The log-Gabor filters are used, as natural textures often exhibit a linearly decreasing log power spectrum. In the frequency domain, the log-Gabor filter bank [12, 13] is defined as

$$G_{ij}(\omega_r, \omega_\varphi) = G(\omega_r - \omega_{r_i^o}, \omega_{\varphi_j^o}) \qquad (1)$$

where (r, φ) are polar coordinates, $\omega_{r_i^o}$ is the logarithm of the center frequency at scale $i \in [1, M_G]$, $\omega_{\varphi_j^o}$ is the jth orientation $(j \in [1, N_G])$ and $G_{\omega_r\omega_\varphi}$ is defined as

$$G_{\omega_r\omega_\varphi} = \exp\left(\frac{-\omega_r^2}{2\sigma_{r_i}^2}\right) \exp\left(\frac{-\omega_\varphi^2}{2\sigma_{\varphi_j}^2}\right) \qquad (2)$$

where σ_{r_i} and σ_{φ_j} are the parameters of the Gaussian. The N_G orientations are taken equidistant (3) and the scales are obtained by dividing the frequency range $\omega_{\max} - \omega_{\min}$

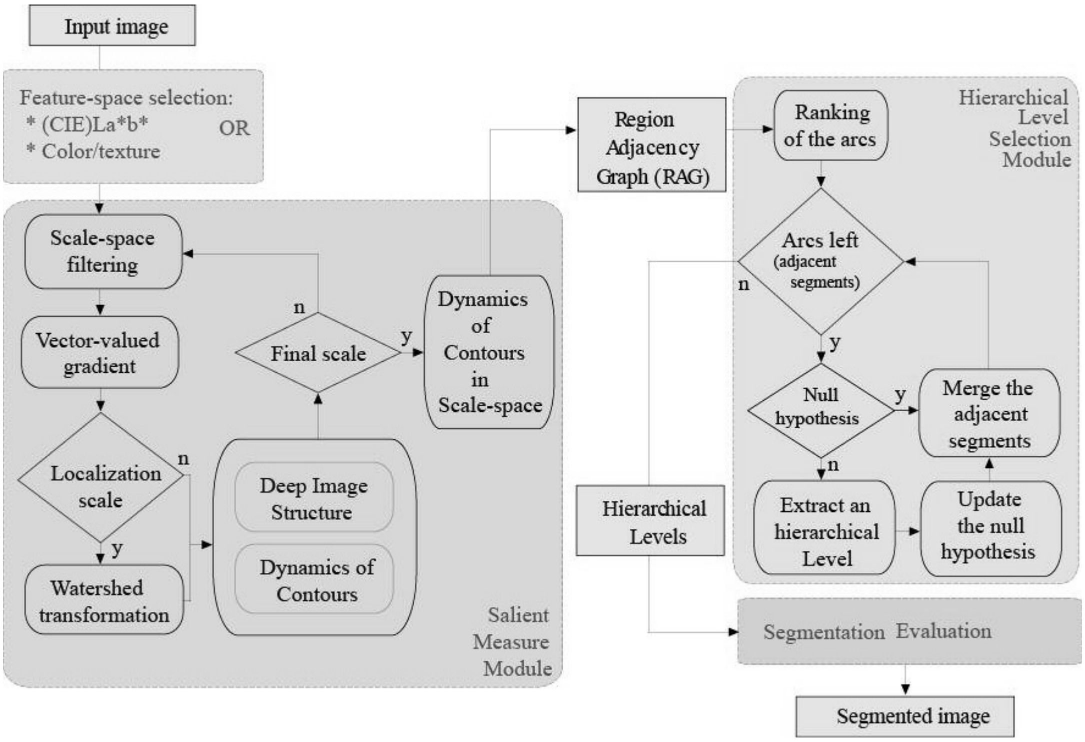


Fig. 1 *Schematic diagram for the automatic multi-scale segmentation scheme for vector-valued images*

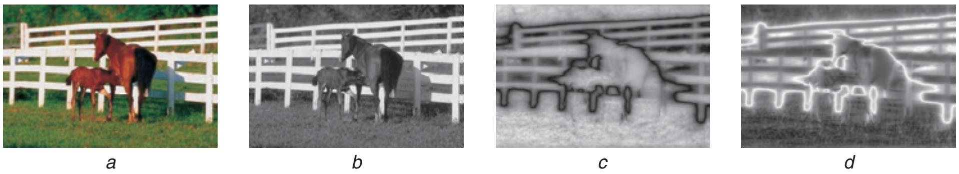


Fig. 2 *Vector normalised texture features*

a Original image
b Luminance component
c Low-pass component
d Textured component

into M_G octaves (4).

$$\sigma_{\varphi_j} = \frac{\pi}{2N_G}$$
$$\omega^0_{\varphi_j} = 2\sigma_{\varphi_j}(j-1) \tag{3}$$

$$\sigma_{r_i} = 2^{i-1}\sigma$$
$$\omega^0_{r_i} = \omega_{\min} + (1 + 3(2^{i-1} - 1))\sigma \tag{4}$$

where $\sigma = (\omega_{\max} - \omega_{\min}/2(2^{M_G} - 1))$ that yields M_G octaves 2σ, $4\sigma, \ldots, 2^{M_G}\sigma$. Note that the maximum frequency cannot be larger than the Nyquist frequency and the dynamics of contours (DC)-component of the image is removed before filtering. We apply the log-Gabor filter on the luminance component (Fig. 2*b*) of the colour image (Fig. 2*a*) to extract the raw texture features.

$$W_{ij} = g_{ij} \otimes L \tag{5}$$

where g_{ij} is the G_{ij} counterpart for the spatial domain, L is the luminance component for which the DC component is removed and $\otimes$ denotes the convolution.

The employed Gabor filter bank consists of one scale and four orientations for which the magnitudes of the responses encode the energy content of the texture feature (Fig. 3*g*–*j*). These feature bands form a hypersphere, in which each vector is further normalised to be a unit vector to emphasise the texture structure information and to reduce any dependencies from the lighting responses.

As mentioned earlier, the segmentation is applied on a colour–texture feature space. However, the inclusion of texture features increases the dimension of the feature-space and hence the computational cost of the segmentation process is increased. Therefore we attempt to determine whether the image contains a sufficient amount of texture to justify the added computational cost. For this, we compute the corresponding low-pass component to identify the non-textured areas. The low-pass component is a Gaussian for which the kernel size is a function of $\omega_{\min}(\sigma_{\text{Lowpass}} = (\omega_{\min}/2)$. It is performed on the image without its DC component. In this work, we used $\sigma_{\text{Lowpass}} = 2$. To determine whether or not the image contains a sufficient amount of textured areas, we compare the average response in the low-pass and the texture components. In the case that the average response of the texture component is lower, we consider only the colour-image in the (CIE)La*b* colourspace, which has the advantage of being perceptual uniform. Otherwise, we create a colour–texture feature space by creating a hypersphere that contains the colour channels and the estimated texture features.

3.1.2 Salient measure module: The main goal of this module is to create a hierarchy among the gradient watersheds detected at the finest scale: the localisation scale. For this purpose, we create at the localisation scale, a RAG, where the nodes represent the detected gradient watersheds and the arcs represent the contours between two watershed segments, that is the adjacencies. To each contour, we attribute a saliency measure comprising the scale-space lifetime (SSL) and the DC in scale-space (DCS) (8) [14, 15]. The entire process to retrieve the saliency measure for the gradient watersheds requires three steps: (i) nonlinear diffusion filtering for creating a scale-space stack; (ii) deep image structure analysis, relating the contours and regions detected at the different scales: at each scale the gradient magnitude of the image is estimated. For successive scales, the duality between the regional minima of the gradient and the catchment basins of the watershed is exploited to make a robust region-based parent–child linking scheme; (iii) contour valuation by

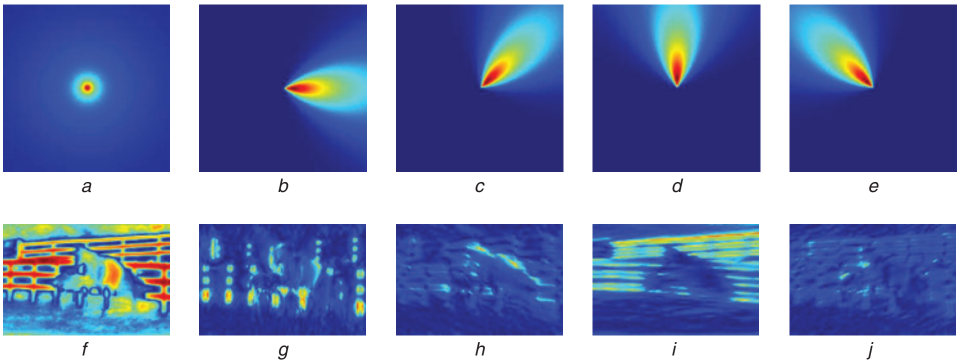


Fig. 3 *Gabor filter bank*

a Low-pass component
b–*e* Gabor filters at the different orientations
f Filter response of low-pass component
g–*j* Corresponding responses of Gabor filters at different orientations

downward projection: the DCS [14, 15] is used to valuate the contours detected at the localisation scale. The latter requires two types of information: (a) the DC [16] at each scale and (b) the deep image structure or scale-space. An overview of the three steps is given subsequently.

Nonlinear diffusion filtering: Scale-space filtering concerns the mechanism that embeds the image into a one-parameter family of derived images for which the image content is causally simplified [17, 18]. The parameter describes the scale or resolution at which the image is represented. The key idea is that important image features persist in scale. In order to avoid blurring and delocalisation of the image features, an image adaptive scale-space filter is used. In this work, we opted for a method that guides the filtering process in such a way that intra-region smoothing, where edges are gradually enhanced, is preferred over inter-region smoothing [11, 19–21]. The employed filter belongs to the class of nonlinear anisotropic diffusion filters. It is a backward diffusion filter that can be interpreted as a 'constraint total variation (TV) minimising flow'. Let $\boldsymbol{I} = \{I^{(1)}, I^{(2)}, \dots, I^{(\mathrm{R})}\}$ be a vector-valued image defined on a finite domain Ω. The scale-space image ($\boldsymbol{u}$) is governed by the following system of coupled parabolic partial differential equations [22]

$$\partial_t \boldsymbol{u}^{(r)} = \mathrm{div}\left[g(|\nabla \boldsymbol{u}_\sigma|) \frac{\nabla \boldsymbol{u}^{(r)}}{|\nabla \boldsymbol{u}^{(r)}|} \right] \quad \forall\ r = 1, 2, \dots, R$$

$$\boldsymbol{u}_{(t=0)} = \boldsymbol{I}$$

$$\partial_n \boldsymbol{u} = 0 \quad \text{on } \partial\Omega \tag{6}$$

where $u^{(r)}$ represents the rth image band, t is the continuous scale parameter and σ is the Catté *et al.* [23] regularisation parameter, which ensures the well-posedness of the above system. The edge stopping function g is formulated as:

$$g(|\nabla \boldsymbol{u}_\sigma|) = \frac{1}{1 + (|\nabla \boldsymbol{u}_\sigma|/K)^2} \tag{7}$$

In the case of backward–forward diffusion filtering, the parameter K (contrast parameter) separates the type of diffusion across the edge. For $|\nabla \boldsymbol{u}_\sigma| < K$, the edge is smoothed and for $|\nabla \boldsymbol{u}_\sigma| > K$ the edge is enhanced. In this diffusion scheme, the edge is always enhanced. The maximum amount of enhancement is obtained at $(K/\sqrt{3})$. We estimate K using the cumulative histogram of the regularised gradient magnitude. A discrete version of the scale-image $\boldsymbol{u}$, denoted as $U = \{u_{t_0}, u_{t_1}, \dots, u_{t_N}\}$, is obtained by applying the natural scale-space sampling method [17]. The finest scale u_{t_0}, (localisation scale), is the scale that obtains a maximum noise reduction while retaining all important image features. Currently, the localisation scale is determined empirically.

Deep image structure analysis: The deep image structure uses a robust region based parent–child linking scheme that is based upon the duality between the regional minima of the gradient and the catchment's basins of the watershed. The linking process is applied using the approach proposed by Pratikakis *et al.* [15], in which the linking of the minima in successive scales is applied by using the proximity criterion [17]. The linking process produces a linkage list for all the detected regions at the localisation scale. Inherently, the latter also yields a linkage list for each adjacency (contour) in the localisation scale. An illustration of both linkage lists is given in Fig. 4.

Contour valuation: In the sequel, we will introduce the reader to the concept of 'DCS' [14, 15], which has been used to valuate the contours detected at the localisation scale. Let $L(a_i) = \{a_i^{(t_0)}, a_i^{(t_1)}, \dots, a_i^{(t_a)}\}$ be the linkage list for the contour a_i, where t_{o} is the localisation scale and the scale t_{a} is the annihilation scale, that is, the last scale in which the contour was detected (annihilation scale). The $\mathrm{SSL}(a_i)$ of a contour is given by $t_{\mathrm{a}} - t_{\mathrm{o}}$ and the DCS is defined as

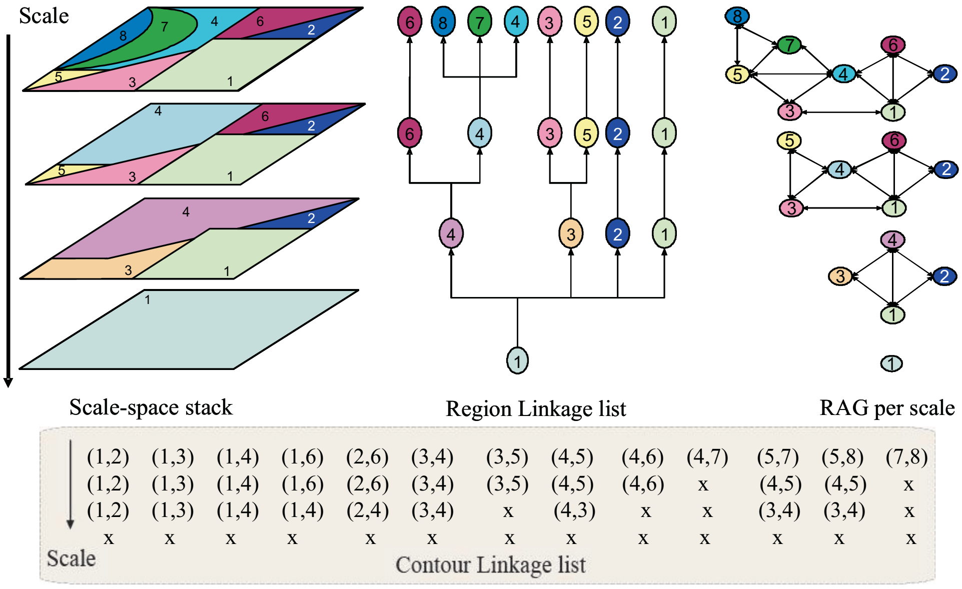
$$\mathrm{DCS}(a_i) = \sum_{b \in L(a_i)} \mathrm{DC}(b) \tag{8}$$

Finally, the saliency measure S attributed to each contour (detected at the localisation scale) is given by

$$S(a_i) = \mathrm{SSL}(a_i) + \frac{\mathrm{DCS}(a_i)}{\max\limits_{\forall b \in A : \mathrm{SSL}(b) = \mathrm{SSL}(a_i)} \mathrm{DCS}(b) + 1} \tag{9}$$

with A being be the set of contours detected at the localisation scale.

In this way, we obtain a hierarchy among the contours, which is consistent with the scale-space filter, that is, the longer a contour persist, in scale the more salient it is. Moreover, the DCS is used to refine the hierarchy among



(1,2)	(1,3)	(1,4)	(1,6)	(2,6)	(3,4)	(3,5)	(4,5)	(4,6)	(4,7)	(5,7)	(5,8)	(7,8)
(1,2)	(1,3)	(1,4)	(1,6)	(2,6)	(3,4)	(3,5)	(4,5)	(4,6)	x	(4,5)	(4,5)	x
(1,2)	(1,3)	(1,4)	(1,4)	(2,4)	(3,4)	x	(4,3)	x	x	(3,4)	(3,4)	x
x	x	x	x	x	x	x	x	x	x	x	x	x

Fig. 4 *Deep image structure*

contours with the same SSL. To find the more salient contour within the set of contours having the same scale-space persistence, we look at the evolution of their contrast in scale-space.

3.1.3 Hierarchical level selection module: This module extracts from the multiscale RAG the different hierarchical levels. It can be seen as some type of global scale-selection method. First, we create a merging sequence by ranking the contours (detected at the localisation scale) according to the saliency measure S. Next, we start merging two successive regions, sharing a common contour, following a hypothesis test that is constructed around a colour similarity measure. For our problem, the hypothesis test is defined as:

- H_0^L: two adjacent regions at level L belong to the same region.
- H_1^L: two adjacent regions at level L belong to different regions.

where H_0^L represents the null-hypothesis and H_1^L denotes the alternative hypothesis. A failure to meet H_0^L indicates that merging the segments under consideration alters significantly the image content significantly according to the current level or scale. Hence, when this occurs, a hierarchical level is extracted and the hypothesis test is updated.

3.1.4 Segmentation evaluation module: For further processing, the extraction of the most suitable hierarchical level is required. We employ a criterion based on a measure that yields a global evaluation of the contrast between the regions and the region uniformity, namely the contrast-homogeneity criterion (CH) [24]. It rewards uniform segments that differ from neighbouring segments. It is formulated by

$$\mathrm{CH}(\boldsymbol{I}, P^L) = \frac{1}{n} \sum_{s_i^L \in P^L} n_i^L \mathrm{CH}_i^L \tag{10}$$

where n is the total number of image pixels, P^L represents the partitioning (set of regions) of the hierarchical level L, s_i^L is the ith region of the partitioning P^L, n_i^L is the size of s_i^L and

$$\mathrm{CH}_i^L = \begin{cases} \dfrac{H_i^L}{B_i^L} & \text{if } H_i^L \leq B_i^L \\ 1 & \text{else} \end{cases} \tag{11}$$

with

$$H_i^L = \frac{1}{n_i^L} \sum_{x \in s_i^k} \|x - \boldsymbol{m}_i^L\|$$
$$B_i^L = \frac{\sum_{j \in A_i^k} n_{ij}^L \|\boldsymbol{m}_i^L - \boldsymbol{m}_j^L\|}{\sum_{j \in A_i^k} n_{ij}^L} \tag{12}$$

where $\boldsymbol{m}_i^L$ represents the average feature vector of s_i^L, $\boldsymbol{A}_i^L$ denotes the set of its adjacent regions and, n_{ij}^L the length of the common boundary between s_i^L and s_j^L. The optimal segmentation is given by the partitioning that minimises the function given in (10). To avoid the selection of extremely over-segmented partitionings, we use an upper limit to the amount of required segments. The latter can be deduced from the image size and the task at hand. In this work, we imposed segmentations with a maximum number of 200 regions. Additionally, we added a minimum number of 20 regions.

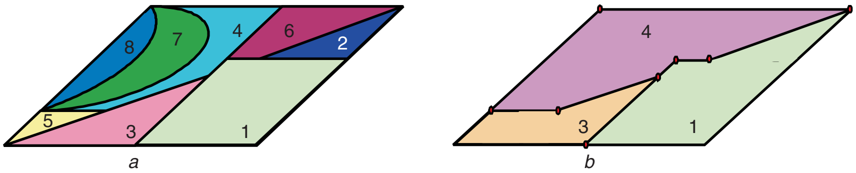


Fig. 5 *Common contour partitioning in the final segmentation*
a Localisation scale
b Final segmentation

3.2 Region features

Having obtained a portioning of the image in significant regions, a set of feature, based mainly on colour, texture and spatial characteristics, will be estimated for each region. We did not use geometric properties, as image segmentation does not always provide a single region for each object in the image, and therefore it is meaningless to compute representative shape features from such regions. The colour space that we use is the RGB colour space. Although, it does not provide the colour compaction of YCrCb and YIQ colour space, neither the perceptual significance of Lab and YUV, our experimental results showed very good performance for retrieval. Other researchers in the area have confirmed our conclusions [7, 25]. Let R_i be a region in the segmented set $\{R\}$ with a set of adjacent regions $\{N(R_i)\}$. In our feature set, we do not only characterise each single region R_i, but we also characterise its neighbourhood by computing relational features. More specifically, the features we compute are described in the following

- Mean colour component

$$\mu \mathrm{C}_k(R_i) = \frac{\sum_{j=1}^{A(R_i)} \mathrm{C}_k(x_j, y_j)}{A(R_i)} \tag{13}$$

- Mean texture component

$$\mu \mathrm{T}_k(R_i) = \int\int |W_k| \mathrm{d}x\, \mathrm{d}y \tag{14}$$

- Variance texture component

$$\sigma^2 \mathrm{T}_k(R_i) = \int\int (|W_k| - \mu \mathrm{T}_k(R_i))^2 \mathrm{d}x\, \mathrm{d}y \tag{15}$$

- Area-weighted adjacent region contrast

$$\mu \mathrm{Con}(R_i) = \frac{\sum_{j=1}^{\mathrm{Card}(N(R_i))} A(R_j) * (\|\mu \mathrm{C}_k(R_i) - \mu \mathrm{C}_k(R_j)\|)}{\sum_{j=1}^{\mathrm{Card}(N(R_i))} A(R_j)} \tag{16}$$

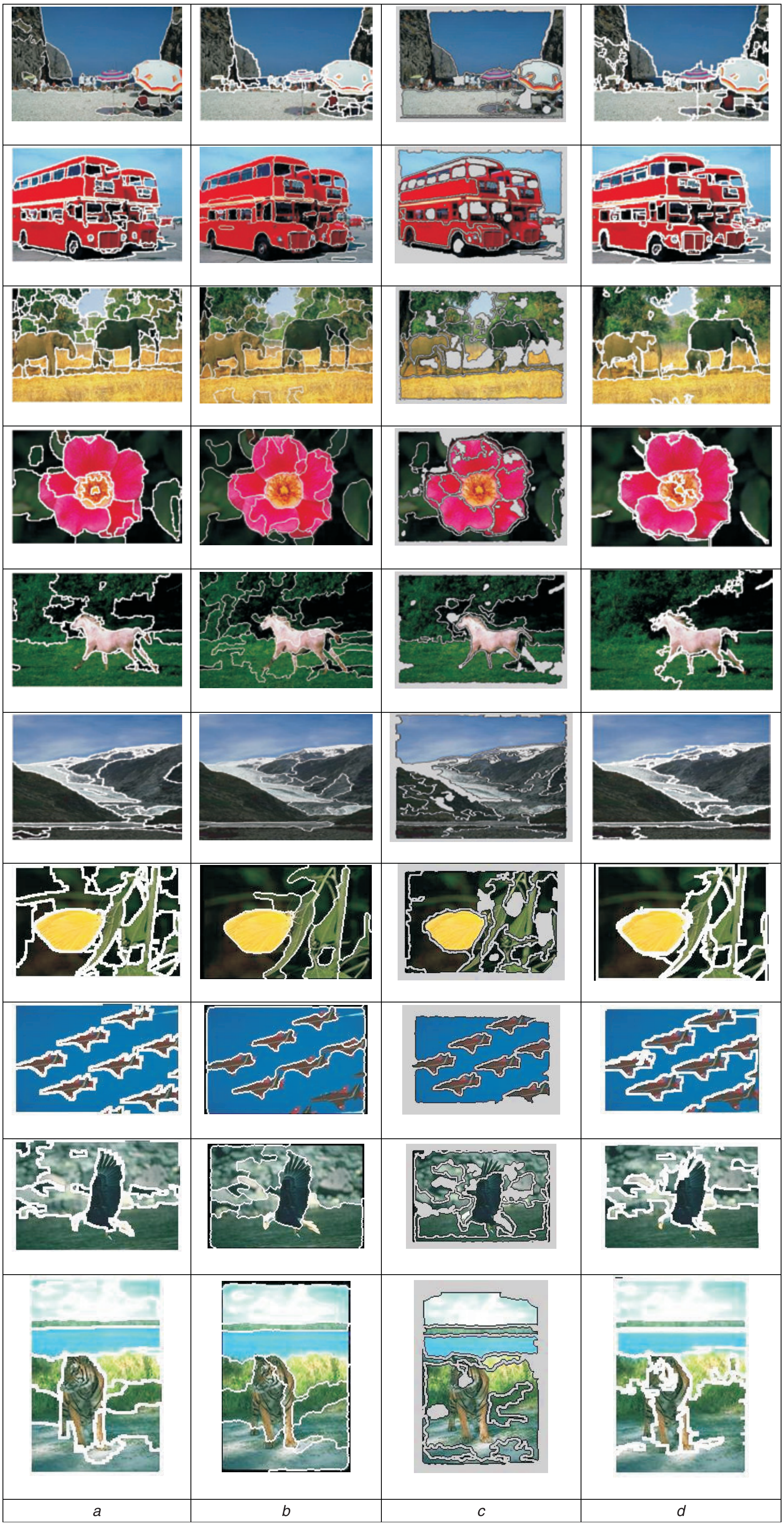


Fig. 6 *Representative segmentation results*

a Using the proposed segmentation scheme
b Using JSEG [27]
c Using E-M algorithm (Blobworld) [2]
d Using graph-based segmentation [28]

- Region geometric centroid

$$\mathrm{G}(R_i; \bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^{A(R_i)} x_i}{A(R_i)}, \frac{\sum_{i=1}^{A(R_i)} y_i}{A(R_i)} \right) \tag{17}$$

where C_k denotes the kth colour component value with $k \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$, T_k denotes the kth texture component value with $k \in [1 \cdot 4]$, $|W_k|$ denotes the magnitude of the transform coefficients of the kth texture component as it is given in (5), $A(R_i)$ denotes the area of region R_i, $\mathrm{Card}(N(R_i))$ denotes cardinality of region's R_i neighbourhood and (x_j, y_j) denotes the coordinates of a pixel that belongs to region R_j.

4 Image retrieval

4.1 Image similarity measure

The EMD [4] is originally introduced as a flexible similarity measure between multidimensional distributions.

Formally, let $Q = \{(\boldsymbol{q}_1, w_{\boldsymbol{q}_1}), (\boldsymbol{q}_2, w_{\boldsymbol{q}_2}), \ldots, (\boldsymbol{q}_m, w_{\boldsymbol{q}_m})\}$ be the query image with m regions and $T = \{(\boldsymbol{t}_1, w_{\boldsymbol{t}_1}), (\boldsymbol{t}_2, w_{\boldsymbol{t}_2}), \ldots, (\boldsymbol{t}_n, w_{\boldsymbol{t}_n})\}$ be another image of the database with n regions, where $\boldsymbol{q}_i$, $\boldsymbol{t}_i$ denote the region feature set and $w_{\boldsymbol{q}_i}$, $w_{\boldsymbol{t}_i}$ denote the corresponding weight of the region. Also, let $d(\boldsymbol{q}_i, \boldsymbol{t}_j)$ be the ground distance between $\boldsymbol{q}_i$ and $\boldsymbol{t}_j$. The EMD between Q and T is then

$$\mathrm{EMD}(Q, T) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}\, d(\boldsymbol{q}_i, \boldsymbol{t}_j)}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{18}$$

where f_{ij} is the optimal admissible flow from $\boldsymbol{q}_i$ to $\boldsymbol{t}_j$ that minimises the numerator of (18) subject to the following constraints

$$\sum_{j=1}^{n} f_{ij} \leq w_{\boldsymbol{q}_i}, \quad \sum_{i=1}^{m} f_{ij} \leq w_{\boldsymbol{t}_j} \tag{19}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{\boldsymbol{q}_i}, \sum_{j=1}^{n} w_{\boldsymbol{t}_j} \right) \tag{20}$$

In the proposed approach, we define the ground distance as follows

$$\begin{aligned} d(\boldsymbol{q}_i, \boldsymbol{t}_j) = \Bigg(& \sum_{k=1}^{3} (\Delta\mu \mathrm{C}_k)^2 + \beta(\Delta\mu \mathrm{Con})^2 \\ & + \sum_{k=1}^{4} (\Delta\mu \mathrm{T}_k)^2 + \sum_{k=1}^{4} (\Delta\sigma^2 \mathrm{T}_k)^2 \\ & + \beta(\Delta \mathrm{G}(i; \bar{x}))^2 + \beta(\Delta \mathrm{G}(i; \bar{y}))^2 \Bigg)^{1/2} \end{aligned} \tag{21}$$

where β is a weighting parameter that enhances the importance of the corresponding features.

4.2 Region weighting

An additional goal during the image retrieval process is to identify and, consequently, to attribute an importance in the regions produced by the segmentation process.

Formally, we have to valuate the weighting factors $w_{\boldsymbol{q}_i}$ and $w_{\boldsymbol{t}_j}$ in (20). Most region-based approaches [3, 6] relate importance with the area size of a region. The larger the area is the more important the region becomes. In our approach, we define an enhanced weighting factor that combines area with scale and global contrast, which can all be expressed by the valuation of DCS (8). More precisely, the weighting factor is computed as follows

$$w_{\boldsymbol{q}_i} = \frac{w_{\mathrm{DCS}_i} * A(R_i)}{\sum_{j=1}^{\mathrm{Card}(R)} w_{\mathrm{DCS}_i} * A(R_i)} \tag{22}$$

$$w_{\mathrm{DCS}_i} = \frac{\sum_{j=1}^{\mathrm{Card}(N(R_i))} (\max \mathrm{DCS}(\alpha_{\mathrm{c}}))}{\mathrm{Card}(N(R_i))} \tag{23}$$

where α_{c} denotes the common border of two adjacent regions at the localisation scale. In (23), we compute the maximum value among the DCS for each adjacency. This occurs because our final partitioning corresponds to a hierarchical segmentation level, wherein a merging process has been applied (Section 3.1.3). Because of merging, any common contour at the final partitioning may contain either a single or a set of contours that correspond to the localisation scale. For the sake of clarity, we refer the readers to Fig. 5, wherein the common contour partitioning in the final segmentation is depicted. More specifically, the common contours in final segmentation (3, 4), (1, 3) and (1, 4) (Fig. 5*b*) consist of the sub-contours set at the localisation scale {(5, 8), (5, 7), (4, 5), (3, 4)}, {(1, 3)} and {(1, 4), (1, 6), (1, 2)}, respectively (Fig. 5*a*).

5 Experimental results

The proposed strategy for CBIR has been evaluated with a general-purpose image database of 1000 images that contain ten categories (100 images per category), taken from the Corel photo galleries [26]. The categories are: 'beaches', 'buses', 'elephants', 'flowers', 'horses', 'mountains', 'butterflies', 'jets', 'eagles' and 'tigers'. Evaluation is performed using precision against recall (P/R) curves. Precision is the ratio of the number of relevant images to the number of retrieved images. Recall is the ratio of the number of relevant images to the total number of relevant images that exist in the database. They are defined as follows

$$\mathrm{Precision}(A) = \frac{R_a}{A} \tag{24}$$

$$\mathrm{Recall}(A) = \frac{R_a}{S} \tag{25}$$

where A denotes the number of images shown to the user (the answer set), S denotes the number of images that belong to the class of the query and R_a denotes the number of relevant matches among A. In our experiments, we have used ten different queries for each category and

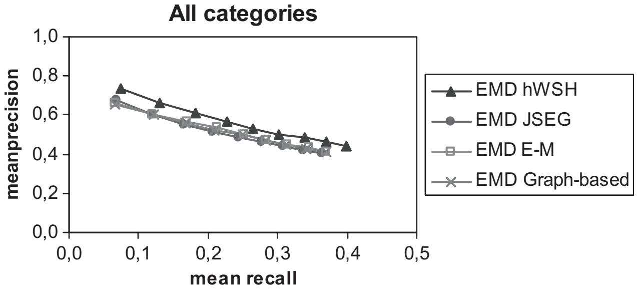


Fig. 7 *Mean precision/mean recall curves over all testing categories*

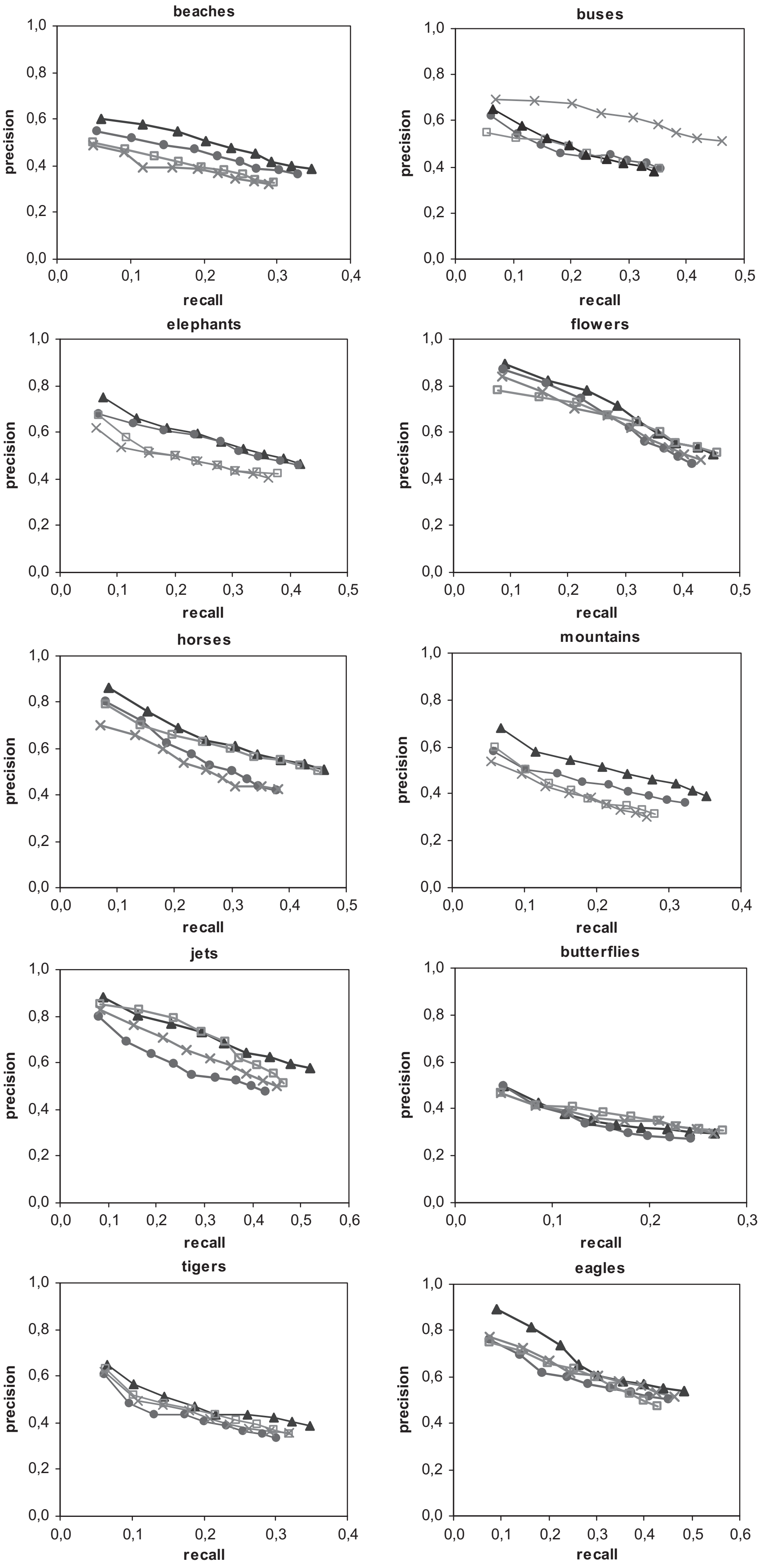


Fig. 8 *P/R curves for individual categories*

we have averaged the P/R values for each answer set. Furthermore, we have used a variety of answer sets that range from 10 to 90 images using a step of ten. For comparison, we have tested our approach, denoted as 'EMD hWSH', with three other region-based image retrieval approaches. Basically, all four approaches differ from each other in the partitioning scheme that they incorporate. The first approach that we compare with 'EMD hWSH' uses the JSEG algorithm [27] for image segmentation. In the presented mean (P/R) curves (Fig. 7), this approach is denoted as 'EMD JSEG'. The second approach that participates in the comparison uses the segmentation approach of the Blobworld CBIR system [2]. This is denoted as 'EMD E-M'. Finally, the third approach uses the graph-based segmentation scheme [28], denoted as 'EMD graph-based'.

First, we compare the segmentation results of the proposed segmentation scheme (hWSH) with the other three schemes. In Fig. 6, we present representative segmentation results of the four segmentation methods, which participate in the comparative study. The multiscale watershed segmentation can capture small but meaningful objects and has often a better localisation of the segment boundaries. Furthermore, it favours slight over-segmentation against under-segmentation because of the selection criterion of the optimal segmentation from the hierarchical levels, which was constructed to penalise under-segmentation. In all cases in which the proposed segmentation scheme produced over-segmented outputs, the other three segmentation algorithms also produced over-segmentations. Apart from the over-segmented exemplars, it is worth-noting that we have achieved excellent segmentations as in the case of categories 'butterflies', 'jets', 'eagles' and 'tigers' where the other schemes produced a mixture of over-segmented and under-segmented results (Fig. 6).

As far as the computational load is concerned, our method is slower compared with JSEG and the graph-based segmentation scheme mainly because of the computational demands for the generation of the multiscale stack. As the involved anisotropic diffusion process of (6) is steered by the image content, its convergence depends on the noise level complexity. It is implemented using the fast numerical scheme proposed by Weickert *et al.* [29]. For vector-valued images, the effort per iteration is proportional to the amount of pixels in the image n and the amount of image channels R. It requires $22nR$ multiplications and divisions, $19nR$ additions and subtraction and nR look-up operations. On average, the creation of the scale-space image needs 150–200 iterations. However, comparing execution times of algorithms, for which the implementations have not been optimised for speed, which is the case for the hWSH, can only give a rough estimate of the execution time.

For each produced region, we compute the feature set that is described in Section 3.2. We would like to note that for 'EMD JSEG', 'EMD E-M' and 'EMD graph-based', we compute region weights by taking into account the area of the region only. We have calculated mean P/R curves over all ten categories (Fig. 7), in which we can observe that 'EMD hWSH' outperforms all other schemes. For the sake of clarity, we provide detailed P/R curves for each individual category in Fig. 8. The individual category analysis shows that in most cases 'EMD hWSH' was the best in performance, whereas there a few cases in which the other schemes were better. Examples are shown in Fig. 8, in which we can observe that the 'EMD graph-based' scheme was clearly the best in category 'buses' as well as that the 'EMD E-M' scheme had a very good behaviour in category 'butterflies'. It is clearly shown that none of the schemes that we compared with had a good behaviour in a consistent way.

6 Conclusions

In this work, we have presented a strategy for unsupervised robust CBIR. The basic components of the proposed scheme are (i) a meaningful watershed-driven hierarchical segmentation that partitions the image into visually consistent homogeneous regions and (ii) a feature set that combines colour, texture and spatial characteristics that are further weighted by a novel weighting scheme that is inherent to the proposed segmentation method. Our experiments have shown that the proposed strategy that does not require any user supervision exhibits a superior behaviour in terms of retrieval accuracy. Considering the computational time of the proposed segmentation scheme, we are working toward faster implementations by considering recursive methods as proposed by Alvarez [30, 31]. Finally, this work can be also used as the initial module in a supervised scheme, in which the user will take into account the resulting initial retrieval. In our future research plans, we plan to exploit such an approach that can further improve retrieval accuracy.

7 References

1 Ma, W., and Manjunath, B.: 'NeTra: a toolbox for navigating large image databases'. Proc. IEEE Int. Conf. Image Processing, 1997, pp. 568–571

2 Carson, C., Belongie, S., Greenspan, H., and Malik, J.: 'Blobworld: image segmentation using E-M and its application to image querying', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, pp. 1026–1038

3 Greenspan, H., Dvir, G., and Rubner, Y.: 'Context-dependent segmentation and matching in image databases', *Comput. Vision Image Understand.*, 2004, **93**, pp. 86–109

4 Rubner, Y., and Tomasi, C.: 'Perceptual metrics for image database navigation' (Kluwer Academic Publishers, Boston, 2000)

5 Fuh, C.-S., Cho, S.-W., and Essig, K.: 'Hierarchical color image region segmentation for content-based image retrieval system', *IEEE Trans. Image Process.*, 2000, **9**, (1), pp. 156–162

6 Wang, J.Z., Li, J., and Wiederhold, G.: 'SIMPLIcity: semantics-sensitive integrated matching for picture libraries', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (9), pp. 947–963

7 Hsieh, J.-W., and Grimson, E.: 'Spatial template extraction for image retrieval by region matching', *IEEE Trans. Image Process.*, 2003, **12**, (11), pp. 1404–1415

8 Mezaris, V., Kompatsiaris, I., and Strintzis, M.: 'Region-based image retrieval using an object ontology and relevance feedback', *EURASIP J. Appl. Signal Process.*, 2004, **9**, (6), pp. 886–901

9 Jing, F., Li, M., Zhang, H.-J., and Zhang, B.: 'An efficient and effective region-based image retrieval framework', *IEEE Trans. Image Process.*, 2004, **13**, (5), pp. 699–709

10 Vanhamel, I., Katartzis, A., and Sahli, H.: 'Hierarchical segmentation via a diffusion scheme in color-texture feature space'. Int. Conf. on Image Processing (ICIP-2003), Barcelona, Spain, 2003

11 Vanhamel, I., Pratikakis, I., and Sahli, H.: 'Multiscale gradient watersheds of color images', *IEEE Trans. Image Process.*, 2003, **12**, (6), pp. 617–626

12 Bigün, J., and du Buf, J.M.: 'N-folded symmetries by complex moments in Gabor space and their application to unsupervised texture segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, **16**, (1), pp. 80–87

13 Rubner, Y., and Tomasi, C.: 'Coalescing texture descriptors' (ARPA Image Understanding Workshop, 1996)

14 Pratikakis, I.: 'Watershed-driven image segmentation', PhD Thesis, Vrije Universiteit Brussel, Brussels, Belgium, 1998

15 Pratikakis, I., Sahli, H., and Cornelis, J.: 'Hierarchical segmentation using dynamics of multiscale gradient watersheds'. 11th Scandinavian Conf. Image Analysis (SCIA 99), 1999, pp. 577–584

16 Najman, L., and Schmitt, M.: 'Geodesic saliency of watershed contours and hierarchical segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (12), pp. 1163–1173

17 Koenderink, J.J.: 'The structure of images', *Biol. Cybern.*, 1984, **50**, pp. 363–370

18 Witkin, A.P.: 'Scale-space filtering'. Int. Joint Conf. Artificial Intelligence, 1983, Vol. 2, pp. 1019–1022

19 Perona, P., and Malik, J.: 'Scale-space and edge detection using anisotropic diffusion', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990, **12**, (7), pp. 629–639

20 Sapiro, G.: 'Geometric partial differential equations and image analysis' (University Press, Cambridge, 2001)

21 Weickert, J.: 'Anisotropic diffusion in image processing', ECMI Series, (Teubner-Verlag, Stuttgart, Germany, 1998)

22 Whitaker, R.T., and Gerig, G.: 'Vector-valued diffusion', in ter Haar Romeny, B.M. (Ed.): 'Geometry-driven diffusion in computer vision' (Springer, 1994), pp. 93–134

23 Catté, F., Lions, P.-L., Morel, J.-M., and Coll, T.: 'Image selective smoothing and edge detection by nonlinear diffusion', *SIAM J. Numer. Anal.*, 1992, **29**, (1), pp. 182–193

24 Vanhamel, I., Pratikakis, I., and Sahli, H.: 'Automatic watershed segmentation of color images', in Goutsias, J., Vincent, L., and Bloomberg, D.S. (Eds.): 'Mathematical morphology and its applications to image and signal processing' computational imaging and vision (Kluwer Academic Press, Parc-Xerox, Palo Alto, CA, USA, 2000), pp. 207–214

25 Gevers, T.: 'Image segmentation and similarity of color-texture objects', *IEEE Trans. Multimedia*, 2002, **4**, (4), pp. 509–516

26 Corel Corp. http://www.corel.com

27 Deng, Y., and Manjunath, B.S.: 'Unsupervised segmentation of color-texture regions in images and video', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (8), pp. 800–810

28 Felzenszwalb, P.F., and Huttenlocher, D.P.: 'Efficient graph-based image segmentation', *Int. J. Comput. Vision*, 2004, **59**, (2), pp. 167–181

29 Weickert, J., ter Haar Romeny, B.M., and Viergever, M.A.: 'Efficient and reliable schemes for nonlinear diffusion filtering', *IEEE Trans. Image Process.*, 1998, **7**, (3), pp. 398–410

30 Alvarez, L., Deriche, R., and Santana, F. (Universidad de Las Palmas de G.C. 'Recursivity and PDE's in image processing'. Technical Report, October 1999, (available online at: http://serdis.dis.ulpgc.es/~lalvarez/research/AlDeSa.ps)

31 Alvarez, L., Deriche, R., and Santana, F.: 'Recursivity and PDEs in image processing'. Proc. Int. Conf. Pattern Recognition, Barcelona, Spain, September 2000, Vol. 1, pp. 242–248