Object-driven content-based image retrieval

Ioannis Pratikakis*, Basilios Gatos and Stavros Perantonis
Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Center for Scientific Research "Demokritos",
153 10 Athens, Greece
E-mail: {ipratika, bgat, sper}@iit.demokritos.gr
*Corresponding author

Iris Vanhamel and Hichem Sahli
Electronics & Informatics Department
Vrije Universiteit Brussel,
1050 Brussels,Belgium
E-mail: {iuvanham, hsahli}@etro.vub.ac.be

Abstract: This paper presents a novel unsupervised strategy for content-based image retrieval. 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.

Keywords: image segmentation, content-based image retrieval

Reference to this paper should be made as follows: Pratikakis, I. Vanhamel, I., Sahli, H. B. Gatos and Sahli H. (2005) 'Object-driven content-based image retrieval', *Int. J. of Signal and Imaging Systems Engineering*, Vol. x, No. x, pp.xx–xx.

Biographical notes: Ioannis Pratikakis received the Diploma degree in Electrical Engineering from the Demokritus University of Thrace, Xanthi, Greece in 1992, and the Ph.D. degree in Applied Sciences from Vrije Universiteit Brussel, Brussels, Belgium, in 1998. From March 1999 to March 2000, he was at IRISA/ViSTA group, Rennes, France as an INRIA postdoctoral fellow. He is currently working as a Research Scientist at the Institute of Informatics and Telecommunications of the NCSR "Demokritos". His research interests include 2D and 3D image analysis, image and volume sequence analysis as well as content-based image / 3D models search and retrieval.

Basilios G. Gatos received his Electrical Engineering Diploma in 1992 and his Ph.D. degree in 1998, both from the Electrical and Computer Engineering Department of Democritus University of Thrace, Xanthi, Greece. He is currently working as a Researcher at the Institute of Informatics and Telecommunications of the National Center for Scientific Research "Demokritos", Athens, Greece. His main research interests are in Image Processing and Document Image Analysis, OCR and Pattern Recognition.

Stavros J. Perantonis is the holder of a BS degree in Physics from the Department of Physics, University of Athens, an M.Sc. degree in Computer Science from the Department of Computer Science, University of Liverpool and a D. Phil. Degree in Computational Physics from the Department of Physics, University of Oxford. Since 1992 he has been with the Institute of Informatics and Telecommunications, NCSRR "Demokritos", where he currently holds the position of Senior Researcher and Head of the Computational Intelligence Laboratory. His main research interests are in Image Processing and Document Image Analysis, OCR and Pattern Recognition.

Iris Vanhamel received the MSc degree in electrotechnical engineering and information processing at the Vrije Universiteit Brussel (VUB) in 1998. She is currently pursuing the PhD degree at the Electronics and Informatics department at VUB. Her research interests include image segmentation, mathematical morphology, scale-space theory and multi-spectral image processing.

Hichem Sahli is currently Professor of image analysis and computer vision with the Department of Electronics and Informatics at Vrije Universiteit Brussel (VUB), Brussels, Belgium. He coordinates the research team in computer vision. His research interests include image analysis and interpretation, computer vision, mathematical morphology, scale-space theory, image registration, image sequence analysis, multispectral image processing.

1 INTRODUCTION

Increasing amounts of imagery due to advances in computer technologies and the advent of World Wide Web (WWW) 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 since it provides the delineation of the objects that are present in an image. Although contemporary algorithms can not 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 content-based image retrieval. 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 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 only user's interaction is the selection of an example image as query. Experimental results demonstrate the performance of the proposed strategy. This paper is organized as follows: Section 2 refers to the image representation along with the proposed feature set which is extracted out of each region. Section 3 is dedicated to the description of the selected similarity metric and a novel region weighting factor while in Section 4 experimental results demonstrate the performance of the proposed CBIR strategy.

2 IMAGE REPRESENTATION

2.1 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 in Vanhamel et al. (2003). It consists of three basic modules. 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***), where the weights incorporate the notion of scale. Using the obtained multiscale RAG, the second module (***Hierarchical Level Selection Module***) extracts a set of partitioning 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.

2.2 Region features

Having obtained a partitioning of the image in significant regions, a set of feature, based mainly on color, texture and spatial characteristics, will be estimated for each region. We did not use geometric properties since 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 color space that we use is the RGB color space. Although, it does not provide the color compaction of YCrCb and YIQ color space, neither the perceptual significance of Lab and YUV, our experimental results showed very good performance for retrieval. 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 characterize each single region R_i but we also characterize its neighborhood by computing relational features. More specifically, the features we compute are described in the following :

- *mean Color component*

$$\mu C_k(R_i) = \frac{\sum_{j=1}^{A(R_i)} C_k(x_j, y_j)}{A(R_i)} \quad (1)$$

- *mean Texture component*

$$\mu T_k(R_i) = \iint |W_k| dxdy \quad (2)$$

- *variance Texture component*

$$\sigma^2 T_k(R_i) = \iint (|W_k| - \mu T_k(R_i))^2 dxdy \quad (3)$$

- *Area-weighted adjacent region contrast*

$$\mu Con(R_i) = \frac{\sum_{j=1}^{Card(N(R_i))} A(R_j) * (\mu C_k(R_i) - \mu C_k(R_j))}{\sum_{j=1}^{Card(N(R_i))} A(R_j)} \quad (4)$$

- *Region geometric centroid*

$$G(R_i; \bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^{A(R_i)} x_i}{A(R_i)}, \frac{\sum_{i=1}^{A(R_i)} y_i}{A(R_i)}\right) \quad (5)$$

where C_k denotes the k^{th} color component value with $k \in \{R, G, B\}$, T_k denotes the k^{th} texture component value with $k \in [1..4]$, $|W_k|$ denotes the magnitude of the transform coefficients of the k^{th} texture component as it is given in Equation (10), $A(R_i)$ denotes the area of Region R_i, $Card(N(R_i))$ denotes cardinality of region's R_i neighborhood and (x_j, y_j) denotes the coordinates of a pixel that belongs to region. For the texture component we use the *log-Gabor* filters since natural textures often exhibit a linearly decreasing log power spectrum. In the frequency domain, the *log-Gabor* filter bank (Bigün and Buf, 1994) is defined as:

$$G_{ij}(\omega_r, \omega_\varphi) = G\left(\omega_r - \omega_{r_i^o}, \omega_{\varphi_j^o}\right) \quad (6)$$

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 j^{th} orientation ($j \in [1, N_G]$) and $G_{\omega_r \omega_\varphi}$ is defined as:

$$G_{\omega_r \omega_\varphi} = \exp\left(\frac{-\omega_r^2}{2\sigma_{r_i}^2}\right)\exp\left(\frac{-\omega_\varphi^2}{2\sigma_{\varphi_j}^2}\right) \quad (7)$$

where σ_{r_i} and σ_{φ_j} are the parameters of the Gaussian. The N_G orientations are taken equidistant Equation (8) and the scales are obtained by dividing the frequency range $\omega_{\max} - \omega_{\min}$ into M_G octaves in Equation (9).

$$\begin{aligned} \sigma_{\varphi_j} &= \frac{\pi}{2N_G} \\ \omega_{\varphi_j}^0 &= 2\sigma_{\varphi_j}(j-1) \end{aligned} \quad (8)$$

$$\begin{aligned} \sigma_{r_i} &= 2^{i-1}\sigma \\ \omega_{r_i}^0 &= \omega_{\min} + \left(1 + 3\left(2^{i-1} - 1\right)\right)\sigma \end{aligned} \quad (9)$$

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

$$W_{ij} = g_{ij} \otimes L \quad (10)$$

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.

3 IMAGE RETRIEVAL

3.1 Image similarity measure

The Earth Mover's Distance (EMD) (Rubner and Tomasi, 2003) is originally introduced as a flexible similarity measure between multidimensional distributions.

Formally, let $Q = \{(\mathbf{q}_1, w_{q_1}), (\mathbf{q}_2, w_{q_2}), \ldots, (\mathbf{q}_m, w_{q_m})\}$ be the query image with m regions and $T = \{(\mathbf{t}_1, w_{t_1}), (\mathbf{t}_2, w_{t_2}), \ldots, (\mathbf{t}_n, w_{t_n})\}$ be another image of the database with n regions, where $\mathbf{q}_i, \mathbf{t}_i$ denote the region feature set and w_{q_i}, w_{t_i} denote the corresponding weight of the region. Also, let $d(\mathbf{q}_i, \mathbf{t}_j)$ be the ground distance between $\mathbf{q}_i$ and $\mathbf{t}_j$. The EMD between Q and T is then:

$$EMD(Q,T) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d(\mathbf{q}_i, \mathbf{t}_j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \quad (11)$$

where f_{ij} is the optimal admissible flow from $\mathbf{q}_i$ to $\mathbf{t}_j$ that minimizes the numerator of Equation (11) subject to the following constraints:

$$\sum_{j=1}^{n} f_{ij} \le w_{q_i}, \sum_{i=1}^{m} f_{ij} \le w_{t_j} \quad (12)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min\left(\sum_{i=1}^{m} w_{q_i}, \sum_{j=1}^{n} w_{t_j}\right) \quad (13)$$

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

$$\begin{aligned} d(\mathbf{q}_i, \mathbf{t}_j) = \; & \Big(\sum_{k=1}^{3}(\Delta\mu C_k)^2 + \beta(\Delta\mu Con)^2 + \\ & \sum_{k=1}^{4}(\Delta\mu T_k)^2 + \sum_{k=1}^{4}(\Delta\sigma^2 T_k)^2 + \\ & \beta(\Delta G(i;\bar{x}))^2 + \beta(\Delta G(i;\bar{y}))^2\Big)^{\frac{1}{2}} \end{aligned} \quad (14)$$

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

3.2 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_{q_i} and w_{t_j} in

Equation (13). Most region-based approaches (Greenspan et al., 2004; Wang et al. 2001) 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 which combines area with scale and global contrast, which can all be expressed by the valuation of dynamics of contours in scale-space (Pratikakis et al., 1999). Let $L(a_i) = \{a_i^{(t_0)}, a_i^{(t_1)}, .., a_i^{(t_a)}\}$ be the linkage list for the contour a_i, where t_o is the localization scale and, the scale t_a is the annihilation scale, i.e. the last scale in which the contour was detected (annihilation scale). The dynamics of contours in scale space (DCS) are defined as:

$$DCS(a_i) = \sum_{b \in L(a_i)} DC(b) \quad (15)$$

More precisely, the weighting factor is computed as follows:

$$w_{q_i} = \frac{w_{DCS_i} * A(R_i)}{\sum_{j=1}^{Card(R)} w_{DCS_i} * A(R_i)} \quad (16)$$

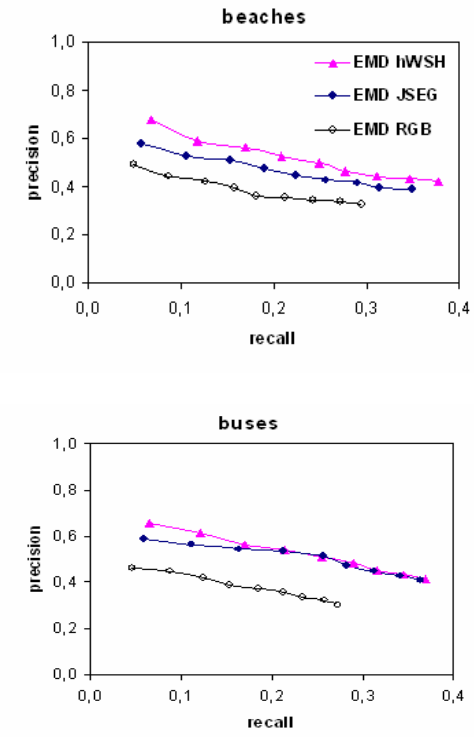
$$w_{DCS_i} = \frac{\sum_{j=1}^{Card(N(R_i))} (\max DCS(\alpha_c))}{Card(N(R_i))} \quad (17)$$

where a_c denotes the common border of two adjacent regions at the localization scale, $A(R_i)$ denotes the area of region R_i and $N(R_i)$ denotes the number of neighbours for region R_i.

4 EXPERIMENTAL RESULTS

The proposed strategy for content-based image retrieval has been evaluated with a general-purpose image database of 600 images from the Corel photo galleries that contain 6 categories (100 images per category). The categories are: “beaches”, “buses”, “elephants”, “flowers”, “horses”, and “mountains”. Evaluation is performed using precision versus 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. To be objective, we have used 10 different queries for each category and we have averaged the precision/recall 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 10. For comparison, we have tested our approach, denoted as "EMD hWSH", with two other region-based image retrieval approaches. All three approaches use as similarity metric the Earth Mover’s Distance (EMD) which is adapted to the underlying feature set of each method. The first approach is based on a k-means clustering (Kanungo et al., 2002) in the RGB color space which feeds the EMD with the produced distributions. In the presented (P/R) curves (Figure 1), this approach is denoted as "EMD RGB". The second approach for comparison that is denoted as "EMD JSEG" uses the state-of-the-art JSEG algorithm (Deng and BManjunath, 2001) for image segmentation.

For each produced region we compute the feature set that is described in Section 2.2. We would like to note that for "EMD JSEG", we compute region weights by taking into account the area of the region only. In the produced P/R curves, we can observe that both "EMD JSEG" and "EMD hWSH" outperform the "EMD RGB". The "EMD JSEG" and "EMD hWSH" methods, have a very good absolute performance after a severe testing of using 10 different queries for each category. This can be attributed to the proposed strategy that is supported by a meaningful proposed feature set along with the proposed similarity metric that both approaches use. Finally, a comparison between "EMD JSEG" and "EMD hWSH" provides a better performance for the proposed scheme ("EMD hWSH”). This can be attributed to a better partitioning that can be achieved using the proposed segmentation scheme compared to JSEG. Considering the overall experimental results, we strongly believe that the proposed strategy for unsupervised image retrieval can guide CBIR applications in a robust way, not only because it can lead to a relatively better performance compared to schemes that other segmentation methods are used but also because our precision / recall curves show that the proposed scheme can achieve an absolute high accuracy.



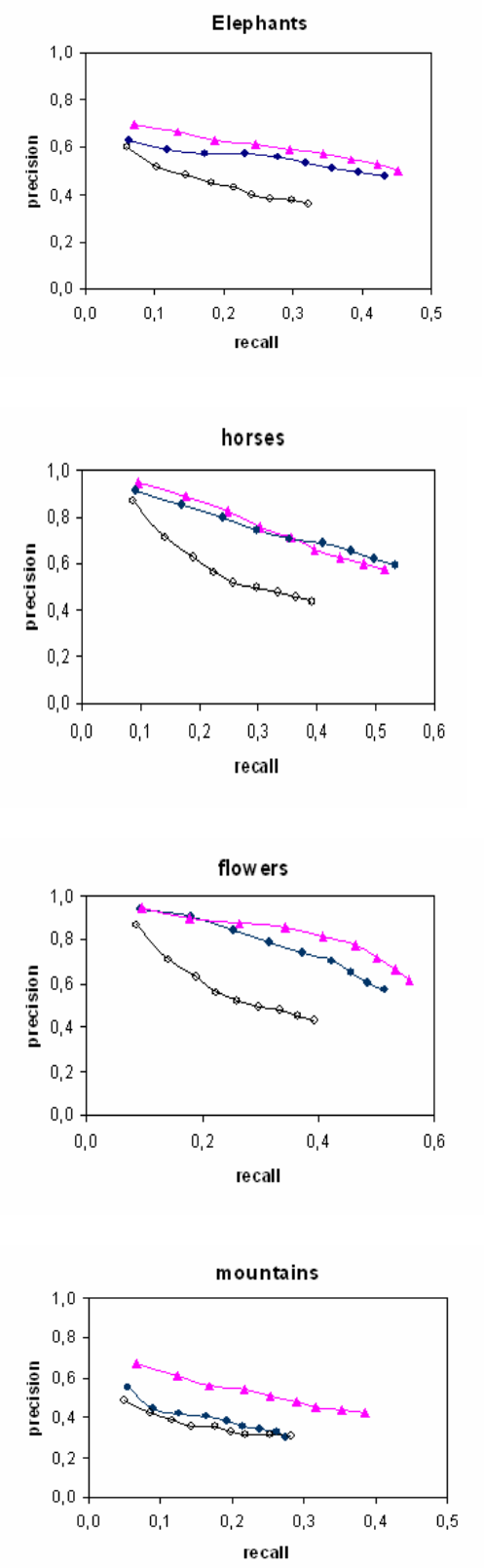


Figure 1: *Precision / recall curves*

REFERENCES

Bigün, J. and du Buf J.M. (1994)'N-folded symmetries by complex moments in Gabor space and their application to unsupervised texture segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 80-87.

Deng, Y. and Manjunath, B.S. (2001)'Unsupervised segmentation of color-texture regions in images and video'. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810.

Greenspan, H., Dvir, G. and Rubner, Y. (2004)'Context-dependent segmentation and matching in image databases', *Computer Vision and Image Understanding*, Vol. 93, pp.86-109.

Kanungo, T., Mount, D., Piatko, C.D., Netanyahu, N.S., Silverman, R., and Wu., A.Y. (2002) 'An efficient k-means clustering algorithm: Analysis and implementation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp.881-892.

Pratikakis, I., Sahli, H., and Cornelis, J. (1999) 'Hierarchical segmentation using dynamics of multiscale gradient watersheds'. In *11th Scandinavian Conference on Image Analysis* (SCIA 99), pages 577–584.

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

Vanhamel, I., Pratikakis, I., and Sahli, H. (2003) 'Multiscale gradient watersheds of color images', *IEEE Transactions on Image Processing*, Vol.12, No. 6, pp.617-626.

Wang, J.,Z., Li, J., and Wiederhold, G. (2001), 'SIMPLIcity: Semantics-Sensitive integrated Matching for picture libraries', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp.947–963.