

A fair ranking method for image database retrieval

L. Costantini^{a,b}, L. Capodiferro^a, M. Carli^b, and A. Neri^b

^a Fondazione Ugo Bordoni, Roma, Italy

^b Applied Electronics Department, University of Roma TRE, Roma, Italy

Abstract. This work aims at organizing the results of query-by-example image database management system so that also the less relevant objects still representing a minority, are presented to the user. Usually, during image retrieval, the objects very similar to the query are ranked in the first positions and shown to the user. On the contrary, objects that slightly differ from the target image, and that are less numerous, are hardly ever shown to the user. The proposed method increases the chances for a *fair* retrieval of multimedia data from digital databases.

Key words: Fair ranking, database extraction, texture analysis.

1 Introduction

Unsupervised or human guided object retrieval is one of the basic functionalities of content-centric Internet services. The first generation of digital database retrieval systems was built on the use of metadata describing the semantic content of a multimedia database, usually extracted by manual procedures. However, Future Internet content aware services require more efficient functionalities for inspection, crawling, recognition, categorization, and indexing of multimedia content with minimal human intervention.

The research activity on Content Based Image Retrieval (CBIR) has been focused on the analysis of local and global image features for the refinement of the coarse annotation, extracted from the WEB pages containing the image, and for re-ranking the results obtained by searching those annotations [1], [3], [4], [5]. In fact, up to the present, traditional search based on keywords has focused on direct use of image content, because of the difficulty in providing at least a sketch of the desired picture. Nevertheless, the widespread diffusion of camera-phones seems to open new opportunities to CBIR for the easiness in providing samples taken from the real world when browsing an image archive by means of a cell phone [2].

Current CBIR techniques are based on comparisons of global features like dominant colors, object shapes, textures as well as of local features relating to the most salient points, like corners. They evaluate a similarity index computed on the basis of the features extracted from the reference template and those of the candidate image. The similarity index can either directly employ these features,

as in the case of adoption, as index, of the maximum likelihood or of the belief functional, or can be based on the comparison of the statistical distribution of the features, as in the case of the use of the Kullback-Leibler divergence. On the other hand, irrespective of the set of local or global features employed for comparing image and video contents, current search engines produce a list of candidates usually sorted with respect to the similarity index. The consequence is that in the presence of several clusters of candidates, the search engine will report the elements of the most similar images and, if the number of elements belonging to that cluster is large enough, the other clusters may be completely cluttered by the dominant one.

The aim of this contribution is to propose a technique for sorting the list of potential candidates in such a way that all the relevant clusters are represented in the list displayed to the user, thus preserving the diversity of possible solutions.

More specifically, a pre-selection of candidates by considering a similarity index accounting only for a subset of the whole feature set is performed. This subset can be either predetermined by referring, for instance, to those features like morphology, that the user regards as more relevant, or automatically selected as those maximizing like a partial similarity index, as described in the next section. Then, the candidate set is clustered by extracting the local maxima of the similarity index evaluated on the full set of features. Finally, for each local maximum, including the absolute one, a representative set of images to be presented as query results is selected. The cardinality of each set may be proportional to the similarity index.

With respect to the evaluation of the partial similarity on the basis of a predefined feature subset, we observe that the relative importance of each feature strictly depends on the user needs. Therefore, although default values for different applications and contests can be specified, the user should have the possibility to provide this information to the search engine.

In the examples reported in this contribution, it is assumed that morphology is the most discriminating global feature for the user, while color based features act as elements discriminating the various clusters. In order to verify the performances of the proposed method in highlighting the differences among the search outcomes, we simplified the set of features employed in the comparison. Thus, with respect to the MPEG-7 ensemble of global and local features, see [1], we focused our attention on the texture morphology, as described by the statistics of the outputs of a set of steerable filters, and on the dominant colors. The simulation results confirm that search results clustering driven by the relative maxima of the similarity index is a powerful technique for preserving web diversity.

2 The fair similarity ranking

Let us model the global feature space V as an M -dimensional vector space, partitioned into L subspaces W_j , each one corresponding to a particular feature set, so that V can be written as direct sum of these subspaces: $V = W_1 \oplus W_2 \oplus$

$\dots \oplus W_M$. Given two elements $\mathbf{x}, \mathbf{y} \in V$, their similarity $s(\mathbf{x}, \mathbf{y})$ can be computed:

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^L s_i(\mathbf{x}_i, \mathbf{y}_i), \quad (1)$$

where $s_i(\mathbf{x}_i, \mathbf{y}_i)$ is the similarity index corresponding to the i -th feature set. Let us consider, without loss of generality, the subset \tilde{V} consisting of N subspaces corresponding to the first N features.

Let $\mathcal{C}_{k,M}$ be the set of k -combinations of the N -element index set, then for any k -combination $\mathbf{i} = (i_1, i_2, \dots, i_{k-1}, i_k)$ let us denote with $W_{\mathbf{i}}$ the set obtained as direct sum of $V_{i_1}, V_{i_2}, \dots, V_{i_{k-1}}, V_{i_k}$, i.e., $W_{\mathbf{i}} = V_{i_1} \oplus V_{i_2} \oplus \dots \oplus V_{i_{k-1}} \oplus V_{i_k}$.

Then given two elements $\mathbf{x}, \mathbf{y} \in V$ we define as partial similarity $\tilde{s}_{k,N}(\mathbf{x}, \mathbf{y})$ based on the best k features out of N as

$$\tilde{s}_{k,N}(\mathbf{x}, \mathbf{y}) = \underset{\mathbf{i} \in \mathcal{C}_{k,N}}{\text{Max}} \sum_{h=1}^k s_{i_h}(\mathbf{x}_{i_h}, \mathbf{y}_{i_h}). \quad (2)$$

We observe that the concept of similarity can be easily replaced by the concept of dissimilarity. Consequently, we define as partial dissimilarity $\tilde{d}_{k,N}(\mathbf{x}, \mathbf{y})$ based on the best k features out of N as

$$\tilde{d}_{k,N}(\mathbf{x}, \mathbf{y}) = \underset{\mathbf{i} \in \mathcal{C}_{k,N}}{\text{min}} \sum_{h=1}^k \tilde{d}_{i_h}(\mathbf{x}_{i_h}, \mathbf{y}_{i_h}). \quad (3)$$

The use of dissimilarity in place of similarity is suggested by the fact that some effective indicators widely used in statistics like the Kullback-Leibler divergence do not satisfy the triangle inequality. Thus, for a given template \mathbf{x} , the set $A(\mathbf{x})$ of candidates can now be computed by thresholding the partial dissimilarity, namely:

$$A(\mathbf{x}) = \{\mathbf{y} \mid \tilde{d}_{k,N}(\mathbf{x}, \mathbf{y}) \leq \gamma\}. \quad (4)$$

Finally, clustering of the candidate set $A(\mathbf{x})$ on the basis of the whole feature space V is performed. The dissimilarity threshold γ_V employed in the cluster formation is, in general, lower than the threshold γ used in the candidate selection phase.

3 Laguerre Gauss decomposition and texture classification

Texture segmentation and classification has been intensively studied and many texture description algorithms have been proposed including statistical [6] and feature based methods [7], [8]. In many applications, segmentation and classification need to be performed without taking object orientations into account, so rotation invariant descriptions of texture patterns are employed.

In this context, a relevant role has been played by the Circular Harmonic Functions (CHF), initially introduced in the field of optical processing for rotation invariant pattern recognition, and recently casted in a wavelet decomposition scheme [10], [11]. Since CHFs are intrinsically tuned to image features like edges, lines, equiangular forks, orthogonal crosses without regard to their orientations, they are good candidates for such applications. A detailed description of Laguerre Gauss wavelet decomposition can be found in [10].

In this paper, we use the pixel classification algorithm employed by Randen and Husoy in their feature set comparisons [9] for segmentation. Then we utilize a multiscale segmentation method, [10], starting from CHF moments. These moments are obtained by the decomposition of a template on complex CHFs that form a complete orthogonal basis on a unit disc. Due to a general property of the CHFs, a pattern can be easily steered by multiplying the expansion coefficients by complex exponential factors whose phase is proportional to the rotation angle, [12], [13]. As a consequence, rotation invariants can be easily obtained by considering the magnitude of the expansion coefficients.

In our method, the texture classification is performed by considering both the structural (morphological) and the color components. To represent a texture pattern morphology, the expansion coefficients of the Laguerre-Gauss transform of the luminance component Y corresponding to a finite set of NK order pairs $\{(n, k) | n = 1, \dots, N, k = 0, \dots, K - 1\}$ are employed. In particular $n = 1, \dots, 3$ and $k = 0, \dots, 3$ at 3 different resolutions have been employed in the examples reported in the following. Since the magnitude of the transform coefficients is rotation invariant, we use their statistics. In particular, as already stated in [10] it is possible to characterize the marginal density of a wavelet decomposition by using a generalized Gaussian function. This distribution is characterized by two parameters, α and β , directly related to the mean and variance distribution. Thus, mean and variance are sufficient to describe the statistical properties of the Laguerre-Gauss transform coefficient magnitudes of a portion of an image. As a consequence, for each point of a given region of interest (ROI) we first compute the KN Laguerre-Gauss transform coefficients. Then, we evaluate the mean and the variance of their magnitudes inside the ROI, so that to each pattern a morphology feature vector of length $2NK$ is associated.

The partial dissimilarity index $\tilde{d}_Y(\mathbf{x}, \mathbf{z})$ associated to the structure of two textures \mathbf{x}, \mathbf{z} is then evaluated as the the Kullback-Leibler distance KLD among the generalized Gaussian probability density functions modeling the statistical behavior of the wavelet coefficient magnitudes:

$$\tilde{s}_Y(\mathbf{x}, \mathbf{z}) = KLD \left[\left| \mathcal{W}_{\mathcal{L}_k^{(n)}} Y_{\mathbf{x}}(\mathbf{b}, 0, a) \right|, \left| \mathcal{W}_{\mathcal{L}_k^{(n)}} Y_{\mathbf{z}}(\mathbf{b}, 0, a) \right| \right]. \quad (5)$$

where $\mathcal{W}_{\mathcal{L}_k^{(n)}} Y_{\mathbf{x}}(\mathbf{b}, 0, a)$ represents the Laguerre-Gauss transform at location \mathbf{b} , rotation φ , and scale a .

In order to evaluate the dissimilarity index related to the chromatic components Cb, Cr we characterize them by their mean and their centered moments

of the second and third order [2], computed as follows:

$$\mu = \frac{1}{N_r N_c} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} p(i, j) \quad (6)$$

$$\sigma = \left[\frac{1}{N_r N_c} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} (p(i, j) - \mu)^2 \right]^{1/2} \quad (7)$$

$$t = \left[\frac{1}{N_r N_c} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} (p(i, j) - \mu)^3 \right]^{1/3} \quad (8)$$

where p represents the pixel chromatic component at location (i, j) , N_r and N_c respectively represent the height and the width of the image. The color information is therefore represented by a features vector of six elements. In this case, the color matching is based on the Euclidean distance among vectors. The partial dissimilarity index based on the chromatic components is therefore computed as follows

$$\tilde{d}_{Cr}(\mathbf{x}, \mathbf{z}) = \left[[\mu_{Cr}(\mathbf{x}) - \mu_{Cr}(\mathbf{z})]^2 + [\sigma_{Cr}(\mathbf{x}) - \sigma_{Cr}(\mathbf{z})]^2 + [t_{Cr}(\mathbf{x}) - t_{Cr}(\mathbf{z})]^2 \right]^{1/2} \quad (9)$$

$$\tilde{d}_{Cb}(\mathbf{x}, \mathbf{z}) = \left[[\mu_{Cb}(\mathbf{x}) - \mu_{Cb}(\mathbf{z})]^2 + [\sigma_{Cb}(\mathbf{x}) - \sigma_{Cb}(\mathbf{z})]^2 + [t_{Cb}(\mathbf{x}) - t_{Cb}(\mathbf{z})]^2 \right]^{1/2} \quad (10)$$

The overall dissimilarity \tilde{D}_{tot} between two textures is finally computed as follows:

$$\tilde{D}_{tot} = \tilde{d}_Y(\mathbf{x}, \mathbf{z}) + \alpha \left[\tilde{d}_{Cr}(\mathbf{x}, \mathbf{z})^2 + \tilde{d}_{Cb}(\mathbf{x}, \mathbf{z})^2 \right]^{1/2} \quad (11)$$

where α is the factor is used to select the relative importance between structural and chromatic components.

4 Clustering algorithm

The clustering algorithm proposed in the follow aims to realize a data ranking starting from a query-by-example image where the images are grouped together in clusters on the basis of a similarity index, while different relevant clusters are preserved and displayed to the user, thus representing at the same time similarity and diversity between the data set. The algorithm consists of four main steps:

1. In the first step the images in the database are ranked according to the similarity to the query image, on the base of Eq. 11. Both luminance and color components are used in this phase. The first 32 images are considered relevant and they will be used in the following steps (see table 4).
2. A new ranking is performed by computing the similarity between the 32 images previously selected with the image ranked the last position of the previous step. The comparison is made by considering both luminance and color components. A new cluster is created with the images presenting a distance value, with respect to the query image, lower than a fixed threshold. These images are removed from the initial cluster. This step is repeated until the original cluster is empty. For each cluster a feature vector is computed by averaging the features of the images belonging to it.
3. The operations performed on the images in the previous step are repeated on the clusters, by operating on their features vectors. This step is iterated until at least one new cluster is created. If the algorithm does not create any new cluster, the threshold is increased and the step is repeated again. This step ends if the threshold becomes higher than a fixed maximum value or the number of cluster is less than a minimum value (6 in our experiments).
4. By rearranging the clusters created at the previous step considering the number of elements it is possible to obtain the final classification, table 4. The image which represents the cluster can be the first one in the cluster or the image with the most similar representative vector to the representative vector of the cluster.

In our experiments, the starting value of the threshold in the step 3 is the same of the threshold used in the step 2. The optimal value is automatically determined during the process by histogram evaluation.

5 Experimental results

The performance of the proposed method are evaluated by using a database composed by 640 images. In the database there are 40 classes of images, each class is composed by 16 images. We analyze the results obtained using each image in the database as query image. The experimental results are shown in table 5. To a better understanding of table 5 it is important to clarify the meaning of *outsider* and *impure cluster*. An outsider is an image that the algorithm puts in a wrong cluster, instead an impure cluster is a cluster which has at least one outsider. The results show that the algorithm creates a right number of cluster, indeed the average number of clusters created is 5.46 and the true average number of clusters is 4.67, furthermore the percentage of impure clusters is very low, only 9.12%.

6 Conclusions

In this contribution a novel method for a *fair* organization of the results of a query-by-example retrieval system is presented. The system is able to show,

Ranking	Image	Ranking	Image
1	Bark.0000.00	17	Bark.0000.02
2	Bark.0000.08	18	Brick.0005.14
3	Bark.0000.09	19	Bark.0000.11
4	Bark.0000.05	20	Wood.0001.01
5	Bark.0000.12	21	Wood.0001.05
6	Bark.0000.01	22	Flowers.0005.13
7	Bark.0000.04	23	Brick.0005.02
8	Bark.0000.13	24	Leaves.00016.15
9	Bark.0000.06	25	Flowers.0005.12
10	Wood.0001.15	25	Wood.0001.09
11	Bark.0000.10	27	Leaves.0016.09
12	Wood.0001.13	28	Brick.0005.01
13	Wood.0001.02	29	Flowers.0005.09
14	Bark.0000.07	30	Leaves.00016.14
15	Wood.0001.12	31	Leaves.00016.13
16	Wood.0001.08	32	Brick.0005.00

Table 1. Ranking of the first 32 images starting from the query image *Bark.0000.00* of the "Vision Texture" database used for the tests.

Ranking	Representant of cluster	Ranking	Representant of cluster
1	Bark.0000.05	5	Leaves.0016.13
2	Wood.0001.01	6	Flowers.0005.09
3	Bark.0000.11	7	Wood.0001.13
4	Brick.0005.14	8	Wood.0001.12

Table 2. Result of the clustering algorithm obtained by using the image *Bark.0000.00*.

Total number of cluster created	3496
Total number of impure cluster created	319
Total number of outsiders	974
Average percentage of impure cluster	9.12%
Average of outsiders in each impure cluster	3.05
Average number of cluster created	5.46
True average number of cluster	4.67

Table 3. Performance results.

together with the most similar match, also some less similar image cluster. In this way, not only objects very similar to the query are shown to the user. On the contrary, objects that slightly differ from the target image, and that are less numerous, are hardly ever shown to the user. The proposed method increases the chances that all the relevant clusters are represented in the list displayed to the user, thus preserving the diversity of possible solutions.

References

1. Weiguang Xu, Yafei Zhang, Jianjiang Lu, Ran Li, Zhenghui Xie, "A Framework of Web Image Search Engine," JCAI, pp.522-525, 2009 Int.Joint Conf. on Artificial Intelligence, 2009.
2. Ruhan He, Kaiming Liu, Naixue Xiong, Yong Zhu, "Garment Image Retrieval on the Web with Ubiquitous Camera-Phone," APSCC, pp.1584-1589, IEEE Asia-Pacific Services Computing Conf., 2008
3. Ottavio Augusto Bizetto Penatti, Ricardo da Silva Torres, "Color Descriptors for Web Image Retrieval: A Comparative Study," SIBGRAPI, pp.163-170, XXI Brazilian Symp. on Computer Graphics and Image Processing, 2008.
4. Yong Zhu, Naixue Xiong, Jong Hyuk Park, Ruhan He, "A Web Image Retrieval Re-ranking Scheme with Cross-Modal Association Rules," pp.83-86, Int. Symp. on Ubiquitous Multimedia Computing, 2008
5. Lin Xie, Yao Zhao, Zhenfeng Zhu, "A Unified System for Web Personal Image Retrieval," IHH-MSP, pp.787-790, Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, 2008
6. T. Quack, U. Mönich, L. Thiele, and B.S. Manjunath, "Cortina: a system for large-scale, content-based web image retrieval," MULTIMEDIA '04: Proceedings of the 12th annual ACM Int. Conf. on Multimedia, pp. 508-511, New York, NY, USA, 2004.
7. K. Stevenson, C. Leung, "Comparative evaluation of Web image search engines for multimedia applications," ICME, IEEE Int. Conf. on Multimedia and Expo, 2005
8. Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Trans. on Image Processing, Vol. 11, No. 2, February 2002
9. Akira Yanagawa, Winston Hsu, Shih-Fu Chang, "Brief Descriptions of Visual Features for Baseline TRECVID Concept Detectors", ADVENT Technical Report 219-2006-5 Columbia University, July 2006
10. A. Neri, G. Jacovitti, "Maximum Likelihood Localization of 2-D Patterns in the Gauss-Laguerre Transform Doman: Theoretic Framework and Preliminary Results" in IEEE Trans. on Image Processing, Vol. 13, No. 1, January 2004.
11. G. Jacovitti, A. Neri, "Multiresolution Circular Harmonic Decomposition", in IEEE Trans. on Image Processing, pp. 3242-3247, Vol.48, n. 11, November 2000.
12. J. Rosen, J. Shamir, "Circular Harmonic Phase Filters for Efficient Rotation - Invariant Pattern Recognition", *Applied Optics*, Vol.27, No.14, pp.2895-2899, July 1988.
13. M.Carli, G. Jacovitti, A. Neri, "Generalized Maximum Likelihood Test for Rotation Invariant Pattern Recognition", *SPIE Conf. Photonics East*, Boston, November 2000.