ORDINAL-BASED SHAPE RETRIEVAL WITH RELEVANCE FEEDBACK

Faouzi Alaya Cheikh, Bogdan Cramariuc and Moncef Gabbouj

Signal Processing Laboratory
Tampere University of Technology
P.O. Box 553, FIN-33101 Tampere, Finland
{faouzi, crama, moncef}@cs.tut.fi

1. ABSTRACT

In this paper we propose to incorporate a feedback loop, into the ordinal correlation framework and apply it to shape-based image retrieval. The user's feedback on the relevance of the retrieval results is used to tune the weights of the similarity measure. Statistics from the features of both relevant and irrelevant items are used to estimate the weights. Moreover, the information accumulated from previous retrieval iterations is used in the weights estimation. A simple measure of the discrimination power is proposed and used to show that the relevance feedback increases the capability of the ordinal correlation scheme to discriminate between relevant and irrelevant objects.

2. INTRODUCTION

The human perception of image similarity is subjective and context-dependent. Moreover, it is very elusive when we try to quantify it using computers and mathematical formulae. Content-based image retrieval (CBIR) systems try to use low-level features such as color, shape and texture to characterize images and estimate their similarity. This similarity is in general estimated as the distance between the feature vectors characterizing two objects or images.

Even though, it is known that low-level features such as edges and textures are essential for our perception of the world around us. Using them in a fully automated way to estimate the similarity between two images or objects, is still not an easy task. The low performance of fully automatic CBIR can be explained mainly by the incapacity of these low-level features to capture the high-level concepts represented in the images and objects. Therefore, interactive CBIR systems are trying to make use of the users' feedback to bridge the gap between low-level features extracted from the images and the high-level concepts a user may be seeking. Moreover, having the user interact with the retrieval process allows to account for the subjectivity of the users' perception.

Relevance feedback was introduced in text retrieval as early as early as the 60's and was shown to improve results significantly. It was shown later that the use of negative feedback could enhance performance strongly [1]. Relevance feedback in content-based image retrieval has been an active research field in the past few years [2], [3]. The methods of relevance feedback are based on the most popular vector model used in information retrieval, and most of the relevance feedback research can be classified into two approaches: query point movement and re-weighting techniques [2].

In the following section we review the ordinal correlation framework we proposed in [4] for shape similarity estimation. And propose in section 4 to use a re-waighting technique to incorporate users' feedback, in the previously mentioned framework. In section 5, some simulation results are presented and conclusions are drawn in section 6.

3. ORDINAL CORRELATION FRAMEWORK

Binary contour images are first transformed using a geodesic distance transform. This transformation allows the dissemination of the information highly concentrated in the single pixel outline of the object into the neighboring region pixels. Which, makes the rest of the correlation estimation less sensitive to small variations of the contour, errors on the alignment and scaling introduced in the first steps of the scheme. The evaluation of image similarity is based on the framework for ordinal-based image correspondence introduced in [4]. Figure 1 gives a general overview of this approach.

Let X and Y be two images, of equal size. In a practical setting, images are resized to a common size. Let $\{X_1, X_2, ..., X_n\}$ and $\{Y_1, Y_2, ..., Y_n\}$ be the pixels of images X and Y, respectively. We select a number of regions $\{R_1, R_2, ..., R_m\}$ and extract the pixels from both images that belong to these regions. Let R_j^X and R_j^Y be the pixels from image X and Y, respectively, which belong to regions R_j, with $j = 1, 2, ..., m$.

The goal is to compare the two transform images using a region-based approach. To this end, we will be comparing R_j^X and R_j^Y for each $j = 1, 2, ..., m$. Thus, each block in image X is compared to the corresponding block in image Y in an ordinal fashion. The ordinal comparison of the two

0-7803-7946-2/03/$17.00 ©2003 IEEE

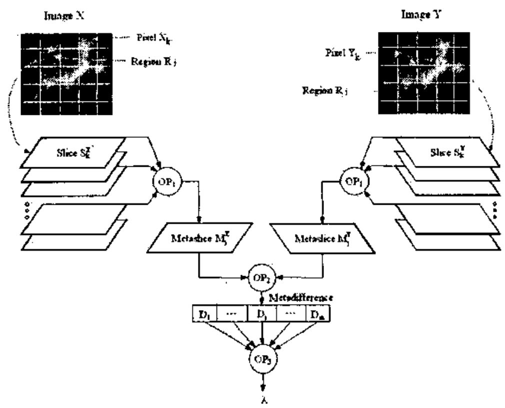

Fig. 1. The general framework for ordinal correlation of images.

regions means that only the ranks of the pixels are utilized. The rank here conveys information on how close a pixel is to the contour. For every pixel X_k, we construct a so-called slice $S_k^X = \{S_{k,l} : l = 1, 2, ..., n\}$, where:

$$S_{k,l}^X = \begin{cases} 1, & \text{if } X_k < X_l \\ 0, & \text{Otherwise.} \end{cases} \quad (1)$$

A slice S_k^X corresponds to the pixel X_k and is a binary image of size equal to image X. Slices are built in a similar manner for image Y as well. The slices corresponding to all the pixels of a region R_j^X are combined via operation $OP_1(.)$ into a metaslice $M_j^X = OP_1(S_k^X : X_k \in R_j^X)$ for $j = 1, 2, ..., m$. Similarly, we combine the slices from image Y to form M_j^Y for $j = 1, 2, ..., m$.

It should be noted that the metaslices are equal in size to the original images and could be multi-valued, depending on the operation $OP_1(.)$. Each metaslice represents the relation between the region it corresponds to and the entire image.

Next a comparison between all pairs of metaslices M_j^X and M_j^Y is performed using operation OP_2, resulting in the meta-difference $D_j = OP_2(M_j^X, M_j^Y)$, for $j = 1, 2, ..., m$. We thus construct a set of meta-difference images $D = \{D1, D2, ..., Dm\}$. The final step is to extract a scalar measure of correspondence from set D, using operation $OP_3(.)$. $\lambda = OP_3(D)$ is called the dissimilarity score between the two objects. It was shown in [5] that this structure can be used to model the well-known Kendall's τ and Spearman's ρ measures [6].

Operation $OP_1(.)$ is chosen to be the component-wise summation operation; that is, metaslice M_j is the summation of all slices corresponding to the pixels in block j or in other words, $M_j = \sum_{k:X_k \in R_j} S_k$. While, operation $OP_2(.)$ is the squared Euclidean distance between corresponding metaslices. That is, $D_j = ||M_j^X - M_j^Y||_2^2$. And operation $OP_3(.)$ sums together all metadifferences to produce $\lambda = \sum_j D_j$, for $j = 1, 2, ..., m$. Small values of λ mean similar objects.

One advantage of this approach over classical ordinal correlation measures is its capability to take into account differences between images at a scale relative to the chosen region size.

4. RELEVANCE FEEDBACK

In the following we propose a re-weighting technique as a straight way of integrating the relevance feedback information into the shape similarity estimation technique we proposed in [4]. This can be easily done by replacing the $OP_3(.)$ of the previous structure by a weighted sum of all the metadifferences to estimate dissimilarity score as $\lambda = \sum_j W_j \times D_j$, where $W_j, j = 1, 2, ..., m$, are the weights of regions R_j, estimated based on the relevance feedback information.

A first search iteration is done in the database using a query image put by the user and weights $W_j = 1$, for $j = 1, 2, ..., m$. Next, the results of this iteration are presented to the user. The user then labels few of these images as relevant and few more as irrelevant according to his information needs.

The goal is to find the weights that would make the dissimilarity scores small for the images in the relevant set, and those of the rest of the images in the database, assumed not relevant to this query. Therefore, making the discrimination between the relevant and irrelevant images easier.

4.1. Weight Estimation

The weights are automatically estimated based on the statistics of each feature separately, without need for user intervention or ad hoc thresholds. Features from both the positive feedback image set and the negative feedback image set are used in the weight estimation. No assumption of optimal relevant set or irrelevant set are made.

For a feature to be useful in discriminating between images:

- its variation among all the images in the database should be large, assuming that the number of relevant images to a certain query is very small compared to the number of irrelevant ones,
- its variation among the relevant images should be small,
- its variation among the irrelevant images should be large.

Since the output of the operation $OP_2(.)$ is a vector containing differences between the metaslices $D_j = ||M_j^X - M_j^Y||_2^2$, for $J = 1, 2, ..., m$. Let Y be the query image and assume that it is the centroid of the relevant set. Therefore, its features can be very well estimated as the mean of the features of the relevant set entries.

Thus, $D_j = ||M_j^X - Mean(M_j^I)||_2^2$, for $J = 1, 2, ..., m$ where I is from the relevant set. It can be seen that the mean of D_j over the relevant set is the squared standard deviation of the M_j^X, for X in the relevant set. Therefore, for a given query image the mean of D_j over a given set of images will be used as an estimate of the variation of the feature vector entries in that set.

We will adopt the following notation:

K: Number of iterative searches with user feedback.

R^0= {all images in the database}.

R^k= {retrieval set after the k'th search}, where

$$k = \begin{cases} 0 & \text{whole database,} \\ 1 & \text{search with the original query,} \\ 2, \ldots, K & \text{searches after user feedback.} \end{cases}$$

R^k_{rel}= {set of relevant images in R^k}. These images are the ones that are marked as relevant by the user at the end of the k'th search, $R^k_{\text{rel}} \subseteq R^k$.

R^k_{irrel}= {set of irrelevant images in R^k}. These images are the ones that are marked as irrelevant by the user at the end of the k'th search, $R^k_{\text{irrel}} \subseteq R^k$, $R^k_{\text{rel}} \cup R^k_{\text{irrel}} \subseteq R^k$.

N: Number of images in the database.

m: Number of features in a feature vector.

f_i= $[f_{i1} f_{i2} \cdots f_{im}]$. Feature vector of the i'th image where f_{ij} is the j'th component of the vector, $i = 1, \ldots, N$, $j = 1, \ldots, m$.

F^k= {feature vectors of the images in R^k} = $\{f_i | i \in R^k\}$.

F^k_{rel}= {feature vectors of the images in R^k_{rel}} = $\{f_i | i \in R^k_{\text{rel}}\}$.

F^k_{irrel}= {feature vectors of the images in R^k_{irrel}} = $\{f_i | i \in R^k_{\text{irrel}}\}$.

F^k_j= {set of values for the j'th components of the feature vectors of the images in R^k} = $\{f_{ij} | i \in R^k\}$.

$F^k_{\text{rel},j}$= {set of values for the j'th components of the feature vectors of the images in R^k_{rel}} = $\{f_{ij} | i \in R^k_{\text{rel}}\}$.

$F^k_{\text{irrel},j}$= {set of values for the j'th components of the feature vectors of the images in R^k_{irrel}} = $\{f_{ij} | i \in R^k_{\text{irrel}}\}$.

Let

$$\mu_j^0 = \text{Mean}(F_j^0), \tag{2}$$

$$\mu^k_{\text{rel},j} = \text{Mean}(F^k_{\text{rel},j}), \tag{3}$$

$$\mu^k_{\text{irrel},j} = \text{Mean}(F^k_{\text{irrel},j}). \tag{4}$$

For the j'th feature in the $k+1$'st retrieval iteration the weight is estimated as

$$w_j^k = \begin{cases} 0, & \text{if } Max(F^k_{\text{rel},j}) \geq Min(F^k_{\text{irrel},j}) \\ \mu^k_{\text{irrel},j} / \mu^k_{\text{rel},j}, & \text{otherwise.} \end{cases} \tag{5}$$

Therefore the weights will be set to zero if the feature (error between relevant image and the query image for a given region) of any of the relevant images is larger than that of any of the irrelevant images, $Max(F^k_{\text{rel},j}) \geq Min(F^k_{\text{irrel},j})$, to avoid increasing the dissimilarity score of the relevant images. Which can simply mean that this region is not relevant to the comparison of the objects, therefore its contribution to the similarity score should be nil. While the weight is computed as the ratio $\frac{\mu^k_{\text{irrel},j}}{\mu^k_{\text{rel},j}}$ which will be larger than one. Therefore, increasing the dissimilarity scores proportionally to the value of the feature F_j^k. This hopefully will increase the dissimilarity scores of the irrelevant images to the query much faster than those of those of the relevant images, making the discrimination easier.

For the case of iterative retrieval, the weights of iteration $k+1$ can either be estimated using all the information accumulated from the previous iterations or just part of it. Using all the history information, $\Omega = \cup_{i=1}^k R^i_{\text{rel}}$, will hinder the convergence of the retrieval process, while using only the previous iteration R^k_{rel} feedback implies that the user has to mark several entries at each iteration to give a good chances for the algorithm to learn his expectations.

We opted for using the feedback info from only two iteration at most for a given search. In this iterative tuning process object relevance decay is implemented, where relevant instances of the near past are considered more heavily than those of the early past. This is done practically in the weight estimation:

$$\mu^{k+2}_{\text{rel},j} = \frac{2 \times \mu^{k+1}_{\text{rel},j} + \mu^k_{\text{rel},j}}{3}, \tag{6}$$

$$\mu^{k+2}_{\text{irrel},j} = \frac{2 \times \mu^{k+1}_{\text{irrel},j} + \mu^k_{\text{irrel},j}}{3}. \tag{7}$$

4.2. Objective Measure of the Discrimination Power

The goal of the weight adaptation is to increase the capability of the similarity measure to discriminate between the relevant and irrelevant images to a given query image. For this purpose we define a simple objective measure of the discrimination power of the similarity estimation scheme. Let $DP_k = \frac{E(\lambda^k_{irrel})}{E(\lambda^k_{rel})}$; where $E(\lambda^k_{rel})$ is the the mean of the similarity scores of the images in R^k_{rel} and $E(\lambda^k_{irrel})$ is the the mean of the similarity scores of the images in R^k_{irrel}.

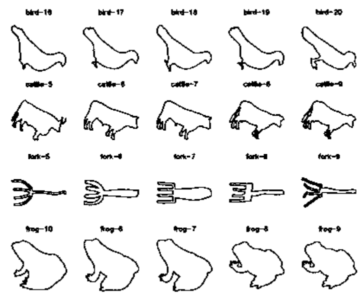

Fig. 2. Contours of test set after alignment.

DP should increase in the second iteration of the search, when the weights are estimated and used in the similarity score computation.

5. EXPERIMENTAL RESULTS

Only four categories of objects were used in this experiment, see Figure 2, and at each iteration three images were used as the relevant set and three as the irrelevant one. The dissimilarity scores of the search results are plotted in Figures 3 and 4 with and without the relevance feedback estimated weights, respectively. It can be noticed from the plot of Figure 4, that the surfaces are flatter on the diagonal of the plot and close to zero, which means that the dissimilarity scores within a group of objects are smaller. And the transitions between the objects of a group and the subjects of another group is much steeper.

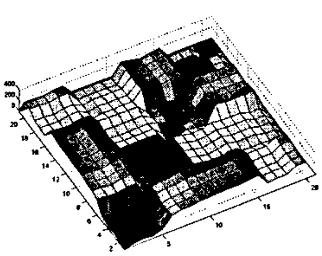

Fig. 3. Similarity scores obtained for the contours in test set presented in Figure 2.

6. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a relevance feedback technique based on re-weighting approach for the ordinal correlation measure, introduced in [4]. We defined an objective measure to evaluate the discrimination power of the modified similarity measure, DP increased from $\simeq 4$ to $\simeq 8$ after the first iteration in our experiment. Based on our experiments we can say that the use of weights increased the performance of the retrieval approach, by better clustering the images of the same category and increasing the distances between this small group of relevant images and those of the rest of the groups. In the future work will focus on optimal weight estimation for a given database.

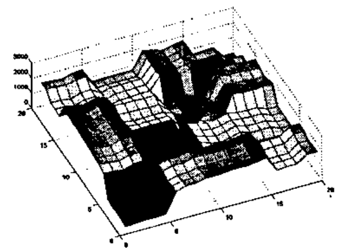

Fig. 4. Similarity scores obtained for the contours in test set presented in Figure 2 using the weights from the feedback images.

7. REFERENCES

[1] J.J. Rocchio. Relevance feedback in information retrieval, in the smart retrieval system. *Experiments in Automatic Document Processing*, pages 313–323, 1971. Prentice Hall, Englewood Cliffs, New Jersey, USA.

[2] Y. Rui, T.S. Huang, S. Mehrotra, and M. Ortega. "Automatic matching tool selection via relevance feedback in mars". In *Proc. of the 2nd Int. Conf. on Visual Information Systems*, pages 109–116, San Diego, California, December 1997.

[3] Aksoy S., Haralick R.M., Alaya Cheikh F., and Gabbouj M. "A weighted distance approach to relevance feedback". In *ICPR*, volume IV, pages 812–815, 2000.

[4] Alaya Cheikh Faouzi, Cramariuc Bogdan, and Moncef Gabbouj. Ordinal-measure based shape correspondence. *JASP*, Special Issue:362–371, April 2002.

[5] Shmulevich I., Cramariuc B., and Gabbouj M. "A framework for ordinal-based image correspondence". In *Proceedings of the X European Signal Processing Conference, EUSIPCO-2000*, pages 1389–1392, Tampere, Finland, September 2000.

[6] Kendall M. and Gibbons J.D. *"Rank Correlation Methods"*. Edward Arnold, New York, 5th edition, 1990.