Shape-based Image Retrieval with Relevance Feedback

Limin Ma, Qiang Zhou, David Chelberg, and Mehmet Celenk
School of Electrical Engineering and Computer Science
Stocker Center, Ohio University, Athens, OH, 45701, USA
{limin.ma,qiang.zhou,chelberg,celenk}@ohio.edu

Abstract
This paper proposes an adaptive framework for shape-based image retrieval with relevance feedback. The motivation is to find an adjustable shape representation scheme that can account for feedback information. Relevance feedback is modeled as a dynamic eigenspace decomposition, and is used to classify the database into relevant and irrelevant groups with respect to the query. Eigenvectors of the subspace are updated by optimizing a linear transform with respect to the $J_3$ class separability criterion. Experimental results show that the proposed approach can effectively capture a user’s perceptual subjectivity.

1. Introduction
Content-based image retrieval (CBIR) has been studied for more than a decade from computer vision perspective. However, early approaches to CBIR only focused on low-level vision and the retrieval performance is still far from satisfactory with respect to human perception. In recent years, the subjectivity of human perception has been identified as one of the most important factors affecting the performance of the content-based image retrieval system. Human perceptual subjectivity refers to the phenomenon that people may perceive the same visual content differently under varying conditions. CBIR with relevance feedback has received much attention in recent years to learn the subjectivity by including users into the retrieval process. Based on different learning mechanisms, a variety of relevance feedback strategies were presented. Rui [1] proposed a heuristic approach to capture the subjectivity by dynamically updating the weights for different features and their associated components. An optimizing learning approach (OPL) presented by Rui and Huang [2] modeled the same phenomenon as an optimization process, in which weights were updated by minimizing the distances between the query and all the relevant results retrieved. However, only positive and labeled samples were used in their case. Discriminant-EM approach [3] formulates the task as a transductive learning problem, in which both the labeled and unlabeled data are used in training. However, the usage of unlabeled data causes significant speed degradation. In the Bayesian approach [4], Gaussian mixture model is adopted as image representation and the Bayesian inference is applied for classification and learning. Tieu and Viola [5] used highly selective features and a boosting technique to learn a classification function in the feature space. Tong and Chang [6] proposed a support vector machine active learning algorithm to select the most-informative images. BiasMap [7] considered the small sample learning issue. An extensive review and comparison of these methods can be found in [8].

Instead of focusing on a general CBIR, our research presented here is restricted to shape-based image retrieval with relevance feedback. Although various shape representation approaches are proposed for MPEG-7 core experiments [9], no single shape descriptor can work well for all cases because of the lack of subjectivity of human perception. We propose an adaptive framework based on eigenspace decomposition and relevance feedback to model the learning part in the retrieval process as an optimal feature selection problem. The novel aspect of our work is that we dynamically change the shape representation to account for the subjectivity of human perception. Shape is projected onto a set of dynamically updated vectorbasis spanning the feature space. Relevance feedback is used to classify the database into relevant and irrelevant groups with respect to the query. Vectorbasis are modified by optimizing a linear transform with respect to the $J_3$ class separability criterion [10]. The major contributions of this paper include 1) providing an adaptive framework for dynamically updating shape descriptors, 2) learning process is modeled as an optimal feature selection problem, and 3) both labeled and unlabeled data are efficiently utilized in the learning process.

The rest of the paper is organized as follows. In section 2, the motivation of this approach is explained. In section 3, the overall algorithms including the framework, shape descriptor, relevance feedback, and adaptation are discussed in detail. Experimental results are presented in section 4. Conclusions and future work are drawn at the end.
2. Motivation

This approach is motivated by the idea that we want to find a feature space in which shapes with similar subjective perception will have the same or similar properties. Shape can be interpreted as a point in a high-dimensional feature space. We expect that in a properly constructed shape representation space, the within-class variance is small while the between-class variance is large as compared to other counter spaces. Shape descriptors such as the Zernike moments, the Fourier series, and wavelets [10] can not fulfill such requirements, even though they define their own spaces in terms of polynomials, complex exponential series, and wavelet packets. Usually, the vectorbasis they use are data-independent. The problem becomes finding a set of data-dependent vectorbasis which is closely related to human perceptual subjectivity. It is well known that principle component analysis (PCA) can derive optimal eigenvectors from samples with respect to a mean square error (MSE) approximation. However, although it is data-dependent and has excellent information reduction properties, it is not necessarily guaranteed in many cases to achieve maximum class separability in the lower dimensional subspaces. Based on these observations, a pattern recognition technique is devised based on $J_3$ criterion to generate optimal eigenvectors. This, in turn, leads to maximum class separability in a subspace in accordance with the classification of perceptual subjectivity.

An experiment illustrates and justifies the validity of this idea. Fig. 1 is given as a comparison, regarding perceptual subjectivity separability, between PCA and $J_3$-based optimal feature selection. Some samples in two subjectively consistent groups and their corresponding reconstruction results based on PCA and $J_3$ optimal eigenvectors are listed in the figure, respectively. It appears that to most human observers the results obtained from the $J_3$ optimal vectorbasis are more "similar" or consistent than that of PCA for different shapes within the same perceptual category. The optimal vectorbasis devised based on the $J_3$ class separability criterion appear to possess more subjective information.

3. Algorithms

3.1. Framework

The framework of this approach is illustrated in Fig. 2. First, a rotation, translation and scale (RTS) invariant shape feature database is created. PCA is applied to the original $m$-dimensional space ($\mathbb{R}^m$) to compute a set of $n$, $(n < m)$, eigenvectors as the initial basis in the $n$-dimensional subspace ($\mathbb{R}^n$). Projection coefficients are treated as shape descriptors. Based on a similarity measurement such as $L_2$ norm, a list of retrieval results is returned for the query. Each result is scored by a user indicating the degree of relevance with respect to the query. In this way, the shape database is classified into two groups; a relevant group and an irrelevant group. This information is implicitly associated with the user’s subjectivity. However, the initial basis is built without such knowledge. The adaptation of shape descriptors involves mapping the original shape features from $\mathbb{R}^m$ to $\mathbb{R}^n$ via a linear transformation and optimizing the transformation with respect to $J_3$ criterion. Thus, a set of vectorbasis with the best discriminatory capability is obtained. This feedback-based adaptation process is carried out iteratively.

3.2. Shape descriptor

In this research, the normalized point list of contours is used as an original shape feature for the sake of simplicity. Translation invariance is achieved by normalizing the contour coordinates with respect to its centroid. Rotation invariance is achieved by rotating the major axis of the shape to the $x$
axis. Scale invariance is satisfied by normalizing the shape with respect to the length of its major axis. The starting point is determined by the intersection between contours and their major axes. Consequently, each shape has two feature vectors. Contours are sampled uniformly at 100 points. PCA is initially employed to generate a set of 13 dimensional vector basis. Projection coefficients are used as shape descriptors.

3.3. Relevance feedback

In this system, at each trial a number of retrieval results are returned based on current basis. Users are asked to score the degree of relevance for each result with respect to the query. The score ranks from 1 to 5, indicating the degree of relevance. The higher the score, the more relevant the result. Score 0 means irrelevant. Naturally, retrieval results are classified into relevant and irrelevant groups. The remaining shapes in the database are classified into the irrelevant group. Finally, the relevance feedback information is utilized and transformed into two distinct classes. As for the scores, they are used to compute the weighted within-class scatter matrix for the relevant group.

3.4. Adaptation

In this module, a new set of vector basis in $\mathbb{R}^n$ needs to be found to separate the relevant and irrelevant classes. Mathematically, the task can be summarized as follows: Transform an $m$-dimensional vector $x$ into $n$-dimensional vector $y$, ($n < m$), $y = A^T x$, so that the class separability criterion is optimized in the $n$-dimensional subspace. A variety of class separability criterion can be chosen such as divergence, the Brattcharyya distance, and scatter matrices-based criteria [10]. For the computational simplicity, the $J_3$ criterion involving within-class scatter matrix ($S_w$) and mixture scatter matrix ($S_m$) is adopted in this paper. $J_3$ is defined in the $y$ subspace as

$$J_3 (A) = trace \{ S_{yw}^{-1} S_{ym} \} = trace \{ (A^T S_{zw} A)^{-1} (A^T S_{zm} A) \}$$

by taking the derivative of $J_3$ with respect to $A$ and setting it to zero, we determine that the optimal transformation matrix $A^*$ must satisfy

$$(S_{zw}^{-1} S_{zm}) A^* = A^* D$$

where $A^*$ is composed of the eigenvectors of $S_{zw}^{-1} S_{zm}$ and the matrix $D$ has the corresponding eigenvalues on its diagonal. By choosing the $n$ largest eigenvectors from $A^*$, a new set of vector basis is obtained, whose class separability is maximized. The detailed definitions and derivations can be found in [10]. In case $S_{zw}$ is singular, a pseudo-inverse is used in place of its inverse matrix [11].

3.5. Computational complexity

With regard to the computational complexity, the major concern involves the update of the scatter matrix and eigenspace decomposition. The reason is that the eigenspace projection and similarity measure simply deal with an inner product operation, which is comparable to OPL [2]. $S_{zm}$ is calculated offline and fixed during the learning process, while $S_{zw}$ has to be updated in each iteration. Fortunately, it can be computed efficiently as $S_{zw} = S_{zm} - S_{zb}$ [10]. Here

$$S_{zb} = \sum P_i (\mu_i - \mu_0) (\mu_i - \mu_0)^T, i = 1, 2$$

where $P_1$, $P_2$ and $\mu_1$, $\mu_2$ are the weights and means of the relevant group ($G_1$) and the irrelevant group ($G_2$), respectively, and $\mu_0$ is the global mean value. The relation between $G_1$ and $G_2$ at steps $t$ and $t + 1$ is a small set of relevant samples ($\Delta G$) moved from $G_2$ to $G_1$; i.e., $G_{t+1} = G_t + \Delta G, i = 1, 2$, assuming that the relevant results returned in previous iterations will still be retrieved in successive iterations. In this way, the mean can be updated as

$$\mu_i^{t+1} = \frac{N_i^t \cdot \mu_i^t + (-1)^{i+1} \cdot N_\Delta \cdot \mu_\Delta}{N_i^t + (-1)^{i+1} N_\Delta}, i = 1, 2$$

here, only the mean ($\mu_\Delta$) and number ($N_\Delta$) of $\Delta G$ need to be computed, while $\mu_1$ and $\mu_2$ can be simply obtained via the above equation. Therefore, the update of scatter matrix is of high efficiency in $O(m^3)$ time. The eigenvectors of $S_{zw}^{-1} S_{zm}$ can be computed by the QR algorithm in $O(m^3)$ time [11].

4. Experimental Results

The experimental database consists of 1100 marine creature images [12]. Ten categories selected by human experts are used as ground truth for this experiment. Each category consists of 30 to 50 images and 10 shapes are randomly selected from each category. Total of 100 queries is made, and the reported retrieval performance is the average of these 100 queries.

The efficiency of the algorithm is evaluated by the convergence rate [1], which indicates how fast the algorithm converges to the user’s true subjectivity. The consistency of a user’s subjectivity is assumed for this experiment. Given a list of $N_{ri}$ retrieval results, the relevance count is defined as

$$count = \sum_{i=1}^{5} n_i$$

where $i$ is the relevance score and $n_i$ is the number of results with score $i$. Suppose $count^*$ is the ideal relevance count and $count (j)$ is the relevance count at the $j$th iteration, the convergence rate ($CR$) is computed at the $j$th iteration as

$$CR = \frac{\text{count^*} - \text{count (j)}}{\text{count^*}}$$
Fig. 3. Convergence rate and \(J_3\)

\[
\text{CR}(j) = \frac{\text{count}(j)}{\text{count}^*}
\]

The ideal case refers to the situation in which all relevant objects are returned. The convergence rate curve with \(N_{rt} = 100\) is shown in Fig. 3a. It clearly shows that the algorithm converges to a nearly ideal case in three iterations. In addition, major increase in \(CR\) is obtained in the first iteration. Successive iterations only contribute to minor increase. The \(J_3\) measure in Fig. 3b shows that the class separability measure is increasing with each iteration.

Precision-recall curve is adopted by most researchers to evaluate the performance of information retrieval system. Precision \((P_r)\) is defined as the number of retrieved relevant objects over the number of the returned objects. Recall \((R_e)\) is defined as the number of retrieved relevant objects over the number of total relevant objects in the database. Three \(P_r\) \((R_e)\) curves corresponding to three iterations are shown in Fig. 4. In the first iteration, \(P_r\) \((R_e)\) curve drops quickly with the increase of recall. In later iterations, high precision is obtained from low to medium recall range, which means that most of the retrieved relevant objects are associated with high rank. In other words, more and more relevant objects are moved from the end of the returned results to the beginning of the returned results.

5. Conclusions

Shape-based image retrieval with relevance feedback is discussed in this paper. An adaptive framework is proposed for finding a feature space, which can optimally interpret the subjectivity of human perception. This approach allows the shape representation to be efficiently modified to account for human perception with respect to the \(J_3\)-based learning criterion. Experimental results show that the proposed method can effectively capture the subjectivity of human perception of shape features. Future work will investigate multiple eigenspace decomposition to extend this approach to multiple perceptual categories.

6. References