A Hierarchical Discriminant Analysis Framework for Content-based Image Retrieval System for Industrial Applications

Kien-Ping Chung, Member, IEEE, and Chun Che Fung, Member, IEEE

School of Information Technology, Murdoch University, Perth, Australia
E-MAIL: k.chung, l.fung@murdoch.edu.au

***Abstract*— Content-based image retrieval (CBIR) systems have drawn interest from many researchers in recent years. One of the potential applications of CBIR is in industrial areas where the most relevant drawings or images can be retrieved speedily without the need to memorize any file name or specific key-words. To increase the retrieval speed, most of the systems pre-process the stored images by extracting a set of predefined features. Such scheme only works well for the server type database systems where the images have been stored previously. It is not feasible for systems that analyze images in real-time where the images are stored or added on an ongoing basis. For instance, personal image search engine for the World-Wide-Web is such an example. In this paper, the authors propose a multi-layer statistical discriminant framework which is able to select the most appropriate features to analyze newly received images thereby improving the retrieval accuracy and efficiency.**

***Index Terms*— Discriminant Analysis, Relevance Feedback, Content-based Image Retrieval System**

I. Introduction

In many industrial applications, an effective image retrieval system is important for the access of the most appropriate images from the database. Examples of such applications are search for specific items among hundreds of inventory items, and, searching of hundreds of images and/or video clips for monitoring and surveillance purposes. While search by keywords or titles can be easily done, the task is difficult if such information is not available. This calls for the need of content-based retrieval systems.

Content-based image retrieval (CBIR) system has been one of the most active areas of research in recent years. One of the main reasons is due to the explosion in the number of images available in digital formats. Combining this factor along with the ease of transferring media through the Internet, digital images become very easy to be obtained by any person for personal, industrial or commercial purposes. CBIR is a system that retrieves an image based on the semantic or visual content of the query which can either be in the forms of key-words or image. The images in these systems are often indexed via key-words, image feature vectors or a combination of both.

One of the issues in handling a large number of images is the need for speedy retrieval. To enhance the retrieval speed, most of the systems often pre-processed the images and stored the extracted features as indexes in the database. Most of the indexes need to be built off line and to be stored in a static manner. However, if one has to retrieve images from the Internet or in some dynamic manner (such as video frames), such approach will not be applicable. On the Internet, there are online-register database systems such as Yahoo or Google. It is however often required to process all the download images at real time. Ideally, to increase the retrieval speed, one has to select an appropriate feature extraction algorithm that best discriminate between the desired (positive) and non-relevant (negative) labeled images. If one is to view the problem: *Given a number of sample data which have been labeled as positive and negative, what features are best to discriminate between the positive and negative labeled data.* The challenge is how to determine the positive or negative labeled images, and how to determine the most appropriate algorithm which is capable to separate the images.

In recent years, relevance feedback has often been used in CBIR systems to interact with users. The goal is to learn the users' intention of their search target. Out of all the approaches [1], *statistical discriminant analysis* and *kernel method* have shown great promises in the ability to analyze the importance of certain features from the given sample data. The ability to classify data given only small amount of sample data with relatively accuracy has intrigued researchers from this field. *Support vector machine* [2, 3] and the *kernel discriminant analysis* [4, 5] are examples of the most popular approaches in recent years. Support vector machine approach has the ability to generalize the common feature characteristics while the other approach has been treated as an optimized learning problem. In particular, discriminant analysis approach has been commonly used in facial recognition systems [5, 6]. The approaches discussed have all shown great potentials. However, all these approaches have treated the input as a collection of flat vectors. In this arrangement, it is rather difficult to determine the discriminant ability of each extracted feature as they all are being treated in the same manner.

Inspired by the discriminant analysis as introduces by Tao and Tang [7] and the hierarchical framework as proposed in the MARS [8] system, this paper proposes a nonparametric discriminant analysis hierarchical relevance feedback framework for the content-based image retrieval system. In addition, the proposed system has the ability to perform automatic feature selection during the retrieval process. This paper first provides a description on the theoretical background, and then followed by the proposed framework. The paper will then discuss the experiment which has been conducted and lastly, a conclusion is drawn based on the findings.

II. Background

A. Discriminant Analysis

Discriminant analysis [9] is a statistical pattern

0-7803-9094-6/05/$20.00 ©2005 IEEE

recognition approach that attempts to maximize the distances between the different labeled data samples. This is achieved by calculating the scatter matrix of the inner- and inter-classes of the different data samples. The scatter matrix is represented in the form of a covariance matrix. As for the inner-class matrix, it is often expressed as:

$$S_x = \sum_{i=1}^{Ny} (x_i - m_x)(x_i - m_x)^T \tag{1}$$

where $\{x_i = 1,\ldots,N_x\}$ denotes the positive examples, and, m represents the mean vector of the positive samples. As for the inter-class matrix, for different types of discriminant analysis approach, there will be corresponding inter-class scatter matrix. The next section provides a description of the inter-class scatter matrix for the nonparametric discriminate analysis.

B. Nonparametric Discriminant Analysis (NDA)

In the nonparametric discriminant analysis approach, the inter-class scatter matrix is derived from the distances obtained from the vectors pointing to the centroid of *another class* of sample data. The main advantage of NDA over other statistical discriminant analysis such as the linear discriminant analysis (LDA) [9] and bias discriminant analysis (BDA) [4] is that NDA does not require all positive samples to be based on a single Gaussian distribution. Hence, NDA does not require an additional kernel transformation matrix to transfer the non-linearly related data to a new feature space for analysis. It therefore eliminates the use of extra parameters. This scatter inter-class matrix for NDA is normally expressed as:

$$S_y = \sum_{i=1}^{Ny} (y_i - m_{yi}^{kx})(y_i - m_{yi}^{kx})^T + \sum_{i=1}^{Nx} (x_i - m_{yi}^{ky})(x_i - m_{yi}^{ky})^T \tag{2}$$

where $\{y_i = 1,\ldots,N_y\}$ are the negative examples being labelled. The inner-class scatter matrix of the NDA is very similar to the other discriminant analysis methods as expressed in (1). It is expressed as:

$$S_x = \sum_{i=1}^{Ny} (x_i - m_x^{kx})(x_i - m_x^{kx})^T \tag{3}$$

where k is the kth nearest sample to the input x_i.

The goal of the discriminant analysis is to find a weight matrix such that the distances between the two scatter class matrixes are maximized. The problem can be expressed as:

$$W_{opt} = \arg\max_{w} \frac{\| W^T S_y W \|}{\| W^T S_x W \|} \tag{4}$$

One can view expression (4) as a problem of generalized eigen-analysis where the optimal eigenvectors associated with the largest eigen-values are the weight factor for the new feature space. By knowing the value of the weights, one can project the new input pattern z onto the new space:

$$Znew\_space = w^T z \tag{5}$$

C. Multi-layer Approach

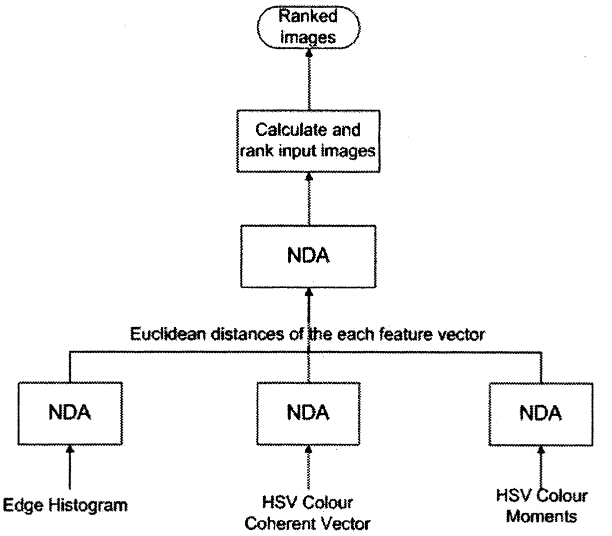

Fig. 1 The proposed HNDA framework.

Fig. 1 illustrates the abstract architectural model of the proposed method. It is essentially based on the proposal by Tao and Tang [7] except that the authors of this paper have restructured the analysis of the input vector into two layers. In this configuration, each image feature vector is processed separately by an individual NDA module. The output of the low-level NDA module is the Euclidean distances of the projected points with reference to the positive centroid. These outcomes serve as inputs to another NDA module which will yield a new point in the final projected space. This new point will be used to compute the final Euclidean distance for ranking purpose. The authors called this configuration as the *Hierarchical Nonparametric Discriminant Analysis (HNDA)* approach.

D. Feature Selection Criteria

As discussed in the previous section, the outcomes from the low-level module shown in Fig. 1 are the Euclidean distances of the projected points with reference to the positive centroid. Hence, the smaller the value indicates the closer the projected point is to the positive centroid. With this information, one can determine the discriminant ability of the individual extracted feature by ranking the distances with the relevancy of the labeled sample data. If the extracted feature is able to discriminate between the two groups of positive and negative data, the distances produced from the positive samples will comprise of the shortest distances within the output as produced by the first layer of NDA modules.

In addition, by applying the discriminant analysis as discussed in Section II, one can expect that the bigger the ratio indicates the better it is to discriminate between the positive and the negative samples. The discriminant ratio can now be shown as follows:

$$ratio = \frac{S_y}{S_x} \tag{6}$$

III. EXPERIMENTAL RESULTS

A. Prototype System

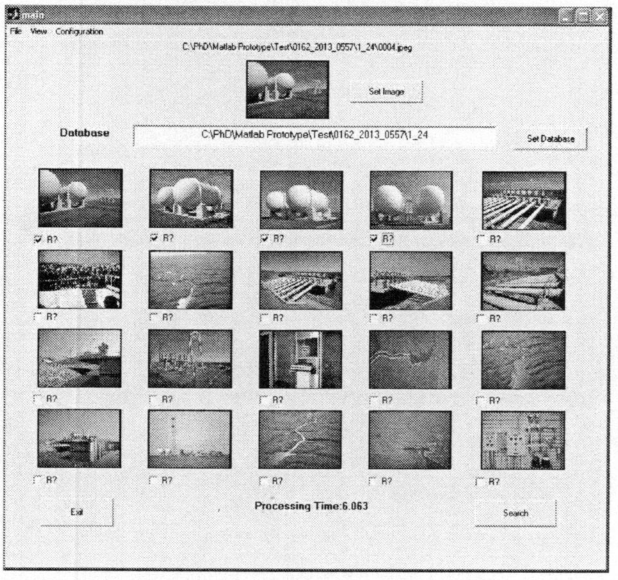

Fig. 2 User interface for the prototype system.

To evaluate the performance of the proposed approach, the authors have designed and implemented a prototype CBIR system using Matlab. The user interface is shown in Fig. 2. This prototype system provides the user with the ability to query the image database with an image sample. After the first retrieval iteration, users can select the relevant images while ignoring the non-relevant images. The system will label the selected images as positive while treating the ignored images as negative. The retrieval procedure of the prototype system is as follows:

1. User inputs a query image.
2. The visual features of the query are extracted by the system.
3. All images in the database are sorted in ascending order based on the distance of dissimilarity.
4. User selects the positive images and the rest will be automatically labelled as negative images.
5. Query the images in the database and project to a transformed space based on the HNDA approach.

For this experiment, the authors have selected the water-filling edge histogram [10], HSV colour coherent vector [11], HSV histogram, global edge direction histogram [12], HSV colour moments [13] and colour intensity histogram. Altogether, there are seven features. Each feature comprises a number of elements. A total number of sixty-six feature elements have been used for this testing. It is the authors' intention to use as many feature elements as possible. Out of the seven features, it is hopeful that at least one or more features will be effective in providing the discrimination between the positive and negative images.

It is difficult to test and verify the retrieval results as the relevancy of an image can be subjective due to different users and purposes. Thus, the authors have developed a set of testing strategies to ensure that the test results can be compared and measured. The idea is to let a user to select the relevant images from a group of images initially. After the user's selection, the system is to re-rank this group of images, and ideally, the user selected images will have a higher rank than the other images. The exact testing steps are as follows:

1. User inputs a query image.
2. User selects an image database.
3. On the first iteration, the system will retrieve and rank the image based on the Euclidean distance measure on the image features. In this iteration, the weights of the features are set as equal.
4. User selects the relevant images from the retrieved images.
5. Base on the current retrieved images, the system will re-calculate the Euclidean distances of each image using the two-layer configuration. The retrieved images are re-ranked accordingly.
6. Accuracy and retrieval speed of the system can be determined from the outputs

B. Results of the Prototype System

The performance of the proposed approach is evaluated according to the retrieval accuracy and distance ratio as defined in (5). The bigger the ratio implies the further the positive images are separated from the negative images, and hence implies better performance. The accuracy of the retrieval is calculated by dividing the number of correctly identified positive images by the number of positive images selected by users.

Concept	Accuracy	Ratio	Feature Selected
Landscape	89%	5.883	3/7
Tank	100%	1109	2/7
Bird	100%	1.755e+007	2/7
Mouse	100%	3975	3/7
Orange Flower	100%	2023	2/7
Yellow Flower	100%	1.715e+003	1/7

Table 1 Retrieval Result from HNDA

Concept	Accuracy	Distance ratio
Landscape	89%	19.12
Tank	100%	4815
Bird	100%	3.8594e+006
Mouse	100%	40
Orange Flower	80%	16
Yellow Flower	100%	1.713e+004

Table 2 Retrieval Result from NDA

Table 1 and Table 2 are the retrieval results gathered from 410 images from the Corel image database. The images were retrieved and classified under six different themes. The result shows that while the retrieval performance of both algorithms is similar, HNDA has the advantage in using fewer features in obtaining the similar retrieval accuracy and in higher speed.

For the above tests, it is observed that similar results have been obtained from the experiment. The accuracy and distance ratio are compatible in both cases. The HNDA approach has given a supreme performance in the cases of retrieving the Bird, Mouse and Orange Flower images. On the other hand, the NDA approach has outperformed in the cases of the Tank and Yellow Flowers. However, it should be noted that the HNDA approach has used only 1 to 3 features out of the total 7. On the other hand, the NDA approach used all 7 features. The retrieval speed for the HNDA approach is roughly proportional to the number of features being used. Hence, HNDA is more efficient than NDA as the latter has to use extra overheads. The situation can be magnified if more images are involved and more resource demanding feature extraction algorithms are used.

IV. CONCLUSION

A new multi-layer framework for the relevance feedback content-based image retrieval system has been introduced in this paper. The proposed framework combines the statistical disciminant approach with multi-layer analysis framework. Using this framework, an improvement has been shown in retrieval speed as compared to the original framework in which the inputs were treated as flat vectors. The test results demonstrate the potential of this framework to be used in systems where only the raw images are stored in the database. Image retrieval from the World-Wide-Web and possible search from inventory databases are typical applications to be benefit from this proposed system.

This paper has shown the testing results for image groups that are visually similar. One of the future directions is to incorporate this framework with images that are similar semantically but differ visually. In addition, this paper only used retrieval accuracy as the feature selection criteria. Other approaches on the incorporation of complex feature extraction algorithms into the feature selection decision making process are also investigated currently.

V. ACKNOWLEDGEMENT

This project is supported by a grant awarded by the 2004 Murdoch University Research Excellence Grant Schemes (REGS). Kien Ping is also a recipient of the Murdoch Research Scholarship (MURS) from 2005.

VI. REFERENCE

[1] X. S. Zhou and T. S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," *ACM Multimedia Systems*, vol. 8, pp. 536-544, 2003.

[2] G.-D. Guo, A. K. Jain, W.-Y. Ma, and H. J. Zhang, "Learning Similarity Measure for Natural Image Retrieval With Relevance Feedback," *IEEE Transactions on Neural Networks*, vol. 13, pp. 811-820, 2002.

[3] S. Tong and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," Proceedings of 9th ACM International Multimedia Conference, Ottawa, Cananda, September 30 - October 05 2001.

[4] X. S. Zhou and T. S. Huang, "Small Sample Learning During Multimedia Retrieval using BiasMap," IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, United States, December 2001.

[5] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms," *IEEE Transactions on Neural Networks*, vol. 14, pp. 117 - 126, 2003.

[6] L.-F. Chen, H.-Y. m. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small smaple size problem," *Pattern Recognition*, vol. 33, pp. 1713 - 1726, 2000.

[7] D. Tao and X. Tang, "Nonparametric Discriminant Analysis in Relevance for Content-based Image Retrieval," Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, United Kingdom, August 23 - 26 2004.

[8] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-Based Image Retrieval with Relevance Feedback in MARS," International Conference on Image Processing, Washington, DC, October 26-29 1997.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed: Academic Press, INC, 1990.

[10] X. S. Zhou and T. S. Huang, "Edge-based Structural Features for Content-based Image Retrieval," *Pattern Recognition Letters*, vol. 22, pp. 457 - 468, 2001.

[11] G. Pass, R. Zabih, and J. Miller, "Comparing Images Using Color Coherence Vectors," The 4th ACM international conference on Multimedia, Boston, Massachusetts, United States, November 18-22 1997.

[12] D. K. Park, Y. S. Jeon, and C. S. Won, "Efficient Use of Local Edge Histogram Descriptor," Proceedings of the 2000 ACM workshops on Multimedia, Los Angles, Calfornia, United States 2000.

[13] M. Stricker and M. Orengo, "Similarity of Color Images," Storage and Retrieval for Image and Video Databases III, San Diego/La Jolla, CA, USA, February 5-10 1995.