CMVF: A Novel Dimension Reduction Scheme for Efficient Indexing in A Large Image Database

Jialie Shen[1], Anne H.H. Ngu[2], John Shepherd[1], Du Q. Huynh[3], Quan Z. Sheng[1]

[1] *School of Computer Science & Engineering*
The University of New South Wales
Sydney NSW 2052, Australia
$\{jls,jas,qsheng\}@cse.unsw.edu.au$

[2] *Department of Computer Science*
Southwest Texas State University
San Marcos, 601 University Drive, Texas, USA
angu@swt.edu

[3] *School of Information Technology*
Murdoch University
Murdoch WA 6150, Australia
d.huynh@murdoch.edu.au

1. INTRODUCTION

In recent years, due to the increasing volumes of multimedia data in the World Wide Web, digital library, biomedicine and other applications, efficient content based similarity search in large image databases is gaining considerable research attentions. As a result, various indexing methods known as *Spatial Access Methods (SAMs)* and *metric trees* have been proposed to support this kind of retrieval. The former includes SS-tree, R^+-tree and grid files; the latter includes the vp-tree, mvp-tree , $GNAT$ and M-tree [3].

However, the optimised distance-based access methods currently available for multidimensional indexing in multimedia databases are developed based on two major assumptions: a suitable distance function is known a priori and the dimensionality of the image features is low. Unfortunately, these assumptions do not make the problem substantially easier to solve. For example, it is extremely difficult to define a distance function that accurately mimics human visual perception in image similarity measurement. Also, typical image feature vectors are high-dimensional (dozens of dimensions). The standard approach to reducing the size of feature vectors is Principle Component Analysis (PCA). However, this approach might not always be possible due to the non-linear correlations in the feature vectors.

Motivated by these concerns, we proposed and developed the $CMVF$ (Combining Multi-Visual Features) framework, a fast and robust hybrid method for nonlinear dimensions reduction of composite image features for indexing in large image database[2]. This method incorporates both the PCA and non-linear neural network techniques to reduce the dimensions of feature vectors, so that an optimised access method can be applied.

In this demonstration, we show that with CMVF approach a small but well-discriminating feature vector can be obtained for effective indexing. It allows us to incorporate classification information based on human visual perception into the indexing. In addition, effectiveness of the indexing can be improved significantly with integration of additional image features. In the following sections, we overview the design and system architecture of our CMVF system, and give performance evaluation.

2. SYSTEM OVERVIEW

An effective content-based retrieval system cannot be achieved by considering only a single type of feature such as colour, texture and shape alone. However, creating an index based on a concatenation of feature vectors(e.g., colour, shape and texture) will result in a very high dimensional feature space, rendering all existing indexing methods useless. Also assuming that each type of visual feature contributes equally to the recognition of that image is not supported in human visual perceptron. We need to "fuse" the multiple single feature vectors into a composite feature vector which is low in dimensions and yet preserves all the necessary information for image retrieval. Thus, non-linear dimension reduction (NLDR) method in conjunction with a multidimensional index structure becomes a natural and practical solution. Figure 1 shows the overall architecture of our hybrid method. The different components of the architecture will be covered in detail in this section.

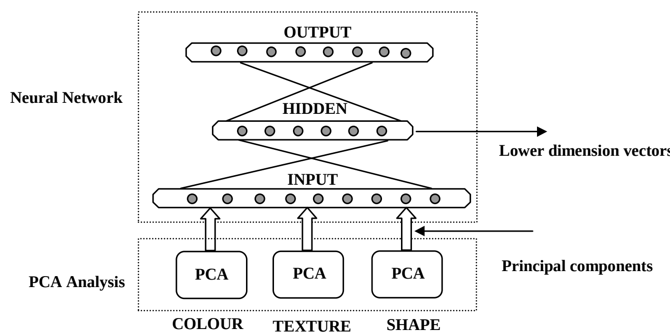

Figure 1: A hybrid image feature dimensions reduction scheme. The linear PCA appears at the bottom, the nonlinear neural network is at the top, and the representation of lower dimensional vectors appears in the hidden layer.

2.1 Composite Image Features

In CMVF, we consider three types of image features: *color*, *texture* and *shape*. Note that our system is not limited to dealing with these three features only. It can be extended to combine other visual and topological features[9] (such as motion and spatial relationship among the objects in the image) for effective indexing.

The colour features are extracted using the colour histogram technique. We used the colour space CIE L*u*v. The reason for selecting the CIE L*u*v instead of the normal RGB or other colour spaces is that it is more uniform perceptually. Our colour features are presented as 37-dimensional vectors.

Texture features carry the property measures, such as *smoothness*, *coarseness* and *regularity*, of an image. The texture features are extracted using a filter-based method. This method detects the global periodicity of intensity values in an image by identifying regions that have high energy, narrow peaks. The advantage of filter-based methods is in their consistent interpretation of feature data over both natural and artificial images. The Gabor filter is a frequently used filter in texture extraction. It measures a set of selected orientations and spatial frequencies. The total number of filters needed for our Gabor filter is 30. Texture features are therefore represented as 30-dimensional vectors.

Shape is an important and powerful attribute for image retrieval. It can represent spatial information that is not presented in color and texture histogram. In our system the shape information of an image is described based on its edges. A histogram of the edge directions is used to represent global information of shape attribute for each image. We used the Canny edge operator[8] to generate edge histogram of images in the prepropressing stage. To solve the scale invariance problem, the histograms are normalized to the number of edge points in each image. In addition, smoothing procedures presented in [1] are used to make histogram invariant to rotation. The histogram of edge directions is represented by 30 bins. Shape features are thus presented as 30-dimensional vectors.

2.2 Architecture of Hybrid Image Feature Dimension Reducer

In CMVF, concatenation[1] is used to form composite feature vectors for further processing. With the 97-dimension feature vectors(37 dimensions for colour, 30 dimensions for texture and 30 dimensions for shape), the PCA[6] is useful as an initial dimension reducer while further dimension reduction for nonlinear correlations can be handled by NLDR. There are two methods for combining the PCA and NLDR:

- Apply the PCA to the single feature vectors separately. The lower-dimensional single feature vectors are then combined to form low-dimensional composite feature vectors for NLDR and classification.
- Apply the PCA to the high-dimensional composite feature vectors. The reduced-dimensional composite feature vectors are then used for NLDR and classification.

[1]Let $\mathbf{x}_c$, $\mathbf{x}_t$ and $\mathbf{x}_s$ be the colour, texture and shape feature vectors, concatenation, denoted by the symbol $\oplus$, of these three feature vectors is defined as follows:

$$\mathbf{x} \equiv \mathbf{x}_c \oplus \mathbf{x}_t \oplus \mathbf{x}_s = \begin{pmatrix} \mathbf{x}_c \\ \mathbf{x}_t \\ \mathbf{x}_s \end{pmatrix}$$

Both methods are adopted in our system so that the differences in the reduction results could be compared.

2.2.1 The PCA for Dimension Reduction

The PCA has been employed to reduce the dimensions of single feature vectors so that an efficient index can be constructed for retrieval in image databases[7]. It has also been applied to image coding, e.g., for removing correlation from highly correlated data such as face images. The advantage of the PCA transformation is that it is linear and that any linear correlations presented in the data are automatically detected. In our system, the PCA is used as a "pre-processing" step in a NLDR method where it provides optimally reduced dimensional feature vectors for the 3-layer neural network, and thus speeds up the NLDR training time.

2.2.2 Classification based on Human Visual Perception

The human perceptual process incorporates information in colour, texture, shape and other visual features under a certain context to classify images into the appropriate classes. To integrate this procedure into our system, we set up a simple on line image classification experiment and asked 7 people (subjects), all of whom are from different backgrounds, to participate in the experiments. Before starting experiment, we first prepared a set of images (labelled as `test-images` from here on), from our 10,000 image collection. This set of image covers all the different classes of images in the collection. In order to enhance robustness of our approach, some of them have image variations(e.g., color distortion, shifting, rotation....etc). At the beginning of each experiment, a query image was arbitrarily chosen from `test-images` and presented to the subjects. The subjects were then asked to pick 20 images that were most similar in colour, texture and shape to the query image. Those images that were selected by more than 3 subjects were classified to the same class as the query image and were then deleted from `test-images`. The experiment was repeated until every image in `test-images` had been categorized into an appropriate class. The end result of the experiments is that images which are similar to each other in colour, texture and shape are put into the same class based on human visual perception. This classification results are used in the NLDR process described below.

2.2.3 Neural Network for Dimension Reduction

In our work, a three-layer perceptron neural network with a quickprop learning algorithm[5] is used to perform dimensions reductions of image features. The network in fact acts as an image classifier. The training samples are training patterns of the form $(\mathbf{v}, c)$ where $\mathbf{v}$ is a feature vector, which can be either a single-feature vector or a composite feature vector, and c is the class number to which the image represented by $\mathbf{v}$ belongs. We note that the class number for each feature vector was determined by the experiments mentioned in the previous subsection.

When the network has been successfully trained, the weights that connect between the input and hidden layers are entries of a transformation that map the feature vectors $\mathbf{v}$ to smaller dimensional vectors. Thus, when a high-dimensional feature vector is passed through the network, its activation values in the hidden units form a lower-dimensional vector. This lower dimensional feature vector keeps the most important information of the original feature vectors (colour, texture and shape).

3. PERFORMANCE EVALUATION

In this section, results from a comparative study are presented to demonstrate superiority of our hybrid dimension reduction method over using the PCA or using neural network alone. We used a collection of 10,000 images. These images were retrieved from different public domains that can be classified under a number of themes which cover natural scenery, architectural buildings, plants, animals, rocks, flags, etc. A subset of this collection of images was selected to form the training samples(Section 2.2.2).

3.1 Performance on Image Categorisation

To determine the accuracy and efficiency of the three methods for dimension reduction, we introduce the measure $class\ separation\ degree$ C_i, defined as:

$$C_i = \frac{\sum_{j=1}^{N} Q_j}{N(M-N)}, \quad i = 1 \dots m$$

where m is the number of classes, N is the number of relevant images. In the class, M is the total number of test images, Q_j is the number of images whose distances to the j-th image in the class are greater than all the distances from the j-th image to its relevant images. An image is said to be $relevant$ to a class if it belongs and has been correctly assigned or classified to that class.

Reduction Method	Average Rate	Feature Vector[2]	Learning Time(epoch[3])
PCA	90.2	$\mathbf{x}_c \oplus \mathbf{x}_t \oplus \mathbf{x}_s$	N/A
NN[4]	100%	$\mathbf{x}_c \oplus \mathbf{x}_t \oplus \mathbf{x}_s$	7100
CMVF	99.9%	$P(\mathbf{x}_c \oplus \mathbf{x}_t \oplus \mathbf{x}_s)$	4200
CMVF	99.9%	$P(\mathbf{x}_c) \oplus P(\mathbf{x}_t) \oplus P(\mathbf{x}_s)$	4120

Table 1: Average class separation values with different method.

From Table 1 it can be seen that all classes of the test image collection are well separated by using neural network and hybrid approach comparing using the PCA. However, dimension reduction with neural network suffers from very long learning time. In contrast, our proposed hybrid method does not lose much accuracy but improve the network learning time. The efficiency is gained by using a relatively small number of network inputs and the network training iterations are conducted in the direction of the largest eigenvalues for each feature.

3.2 Evaluation of Reduced Dimensional Image Features using M-trees

We used M-trees[4] as access method for evaluating the quality of feature space reduced by the PCA, neural network and hybrid method. The number of dimensions of M-trees was set to 6, corresponding to the number of hidden units used in the neural networks. Every image from the collection can serve as a query image. We posed a query image to the M-trees to conduct a k-NN search, where k was set to 100. The concepts of $normalized\ precision$ and $normalized\ recall$[10] in information retrieval were used to evaluate the effectiveness of similarity retrieval since not all relevant images are retrieved.

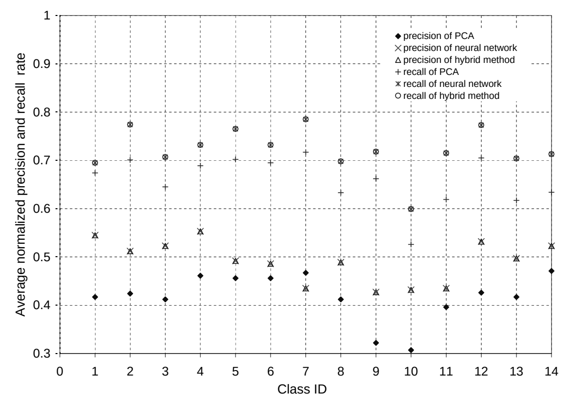

Figure 2: Comparing hybrid method with the PCA and neural network on average normalized recall and precision rate.

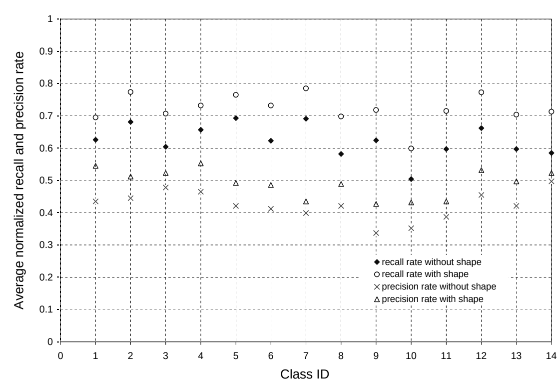

Figure 3: Effectiveness of adding shape feature on hybrid method

In Figure 2, we can see that the $normalized\ recall$ and $normalized\ precision$ values from the neural network and the hybrid methods are almost the same. Thus, the major difference between two approaches is the time required to train the network. One can therefore conclude that it is more advantageous to use a hybrid dimensions reduction method to reduce the dimensions of image features for effective indexing using M-tree. In addition, system performance can be improved considerablely with incorporation of other visual features. As is evident from Figure 3, an addition of shape feature into our system gave approximately 15% improvement of recall and precision over just using color and texture histogram.

---

[2]Because there is no difference in the results of methods used to organise the input feature vectors, we just present one of them.

[3]Epoch means one complete presentation of the input data to the network being trained.

[4]NN means neural network.

3.3 Robustness

Robustness is a very important feature for a content based image retrieval system because image data in real life always accompanies with noise and distortion. With incorporation of human visual perception, CMVF is robust to different kind of image variations including color distortion, sharpness changes, shifting and rotation. Experiment shows that CMVF is robust to 50.4 percent sharpening, 45 degree rotation, blurring with a 9 x 9 Gaussian filter, random spread by 10 pixels, 10 percent more saturation, 11 percent less saturation and pixelization by 9 pixels.

4. DEMONSTRATION

With the use of hybrid structure, CMVF illustrates its great advance in performance against other dimension reduction methods such as the PCA and neural network. To show these advance, a content based image retrieval system has been developed in C++ and Java. An online demonstration is provided[5]. When the user accesses the CMVF webpage, a list of images are randomly selected and displayed as potential query images. The user can submit one of them as a query and the system will search for the images that are most similar in visual content. It displays a list of similar images, in order, starting from the most similar. The query can be executed with any of the following retrieval methods: PCA, neural network, CMVF and CMVF with shape. During this demonstration, we will present its advance in effectiveness, flexibility and robustness via the following:

- Effectiveness: One of our conjectures is that it is possible to obtain effective retrieval from low-dimensional indexing vectors, if these vectors are carefully constructed. In CMVF, we build indexing vectors from high-dimensional "raw" feature vectors via PCA and a trained neural network classifier, which incorporates manual classification criteria. Although some time is required to train the neural network involved in CMVF, we will demonstrate that significant improvement in classification and similarity search can be achieved by CMVF than can the PCA. In comparison with the pure neural network approach, CMVF also gives good classification and query results, with less training time and simpler system structure.

- Flexibility: For further investigations into content-based image retrieval, it would be useful if new indexing features could be easily incorporated into the system. The system demonstrates retrieval based on colour and texture, as well as colour, texture and shape. It was relatively straightforward to incorporate shape into the system, and it clearly demonstrates that the addition of shape leads to superior retrieval results.

- Robustness: In the real world, perfect image data can not be expected. Thus, it is very important for image retrieval systems to be robust to image variations such as color distortion, sharpness changes, shifting and rotation. We will demonstrate that CMVF works effectively even in the presence of the kinds of distortion situations just mentioned.

[5] http://www.cse.unsw.edu.au/~imagedb/MVindex/index.html

5. CONCLUSION

In this demo, we present CMVF, a novel indexing scheme by combining different types of image features to support queries that involve composite multiple features. We have also demonstrated the output quality of our hybrid method for indexing the image collection using M-trees. Our proposed hybrid dimension reduction approach, significantly reduces the size of image feature vectors while at the same time retaining effective discrimination power and also allowing us to incorporate aspects of human visual perception in the weights of the network. This enables any existing access method for moderate dimensions to be used efficiently and effectively.

6. REFERENCES

[1] A.K.Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.

[2] Anne.H.H.Ngu, Q. Z.Sheng, D. Q.Huynh, and R. Lei. Combining multi-visual features for efficient indexing in a large image database. *The VLDB Journal*, 9(4):280–293, May 2001.

[3] C. Böhm, S. Berchtold, and D. A. Keim. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Computing Surveys*, 33(3):322 – 373, September 2001.

[4] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proceedings of the 23rd VLDB International Conference*, pages 426–435, Athens,Greece, September 1997.

[5] S. Fahlman. An empirical study of learning speed for back-propagation networks. Technical Report CMU-CS 88-162, Carnegie-Mellon University, 1988.

[6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[7] G.M.P.Euripdes and C. Faloutsos. Similarity searching in medical image databases. *IEEE Trans. Knowl. Data Eng.*, 3(9):435–447, June 1997.

[8] J.Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, November 1986.

[9] M. Nabil, Anne.H.H.Ngu, and J.Shepherd. Picture similarity retrieval using the 2d projection interval representation. *IEEE Trans. Knowl. Data Eng.*, 8(4):533–539, 1996.

[10] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1993.