

Historic Chinese Architectures Image Retrieval by SVM and Pyramid Histogram of Oriented Gradients Features

¹Bailing Zhang, ²Yonghua Song, ¹Sheng-uei Guan and ³Yanchun Zhang

¹Department of Computer Science and Software Engineering,
Xi'an Jiaotong-liverpool University, Suzhou, 215123, P.R. China

²Department of Electrical Engineering and Electronics,
University of Liverpool, L69 3B, United Kingdom

³School of Computer Science and Mathematics, Victoria University, VIC 3011, Australia

Abstract: Content-Based Image Retrieval (CBIR) of historic Chinese architecture images is an important area of research bridging society, culture and information technology. Most of the image features used in previous content-based image retrieval systems such as colour, texture and some simple shape descriptors are not effective in describing building images due to high variability in the heterogeneous architectural image collections. This study investigates content-based architectural image retrieval mainly by shape features. The recently proposed shape descriptor, Pyramid Histogram of Oriented Gradients (PHOG) features, counts occurrences of gradient orientation in localized portions of an image and has been proved as an efficient tool for providing spatial distribution of edges. Many existing image retrieval systems attempt to compare the query image with every target image in the database to find the top matching images, resulting in an essentially linear search which is prohibitive when the database is large. To solve the problem, it propose to introduce classification as the first stage in the retrieval system. Based on the PHOG features, it apply the Support Vector Machine (SVM) to automatically classify the ancient Chinese architecture images. Cross-validation test results indicate that the generalization performance of the SVM was over 60% compared to neural network's accuracy of 30% and kNN's accuracy 50%.

Key words: Content-based image retrieval, Chinese historical architectures, support vector machine, pyramid histogram of oriented gradient, cross validation, China, Australia

INTRODUCTION

In Content-Based Image Retrieval (CBIR) systems, a set of features such as color, texture and shape (Hiremath and Pujari, 2007; Bowyer and Flynn, 2000; Liu *et al.*, 2007) is used to represent a digital image. Retrieval is performed by comparing the feature vector of the query image against those of the database images using a matching algorithm. Despite over a decade of research into CBIR, the task of finding a desired image in a large collection remains problematic. Even in areas where there is a clear need for effective image retrieval, such as medical and forensic applications, current technology fails to meet user needs. This dilemma can be partly reflected by the fact that most commercial internet search engines still rely on text information to index Web images. Among the various challenges, the selection of an appropriate set of visual descriptors that capture the particular properties of a specific domain and the

distinctive characteristics of each image class remains as the primary concerns. Much existing CBIR research has concentrated on retrieval techniques for natural images (typically photographs of natural scenes or objects) using various combinations of extracted colour, texture and layout feature (Veltkamp *et al.*, 2001). Techniques for the retrieval of certain special architecture images have received less attention even though there is evidence that these images require different techniques for effective retrieval. For the retrieval of historical Chinese architecture images, it is obvious that the colour and texture features are highly variable according to different illumination conditions and viewing positions. On the other hand, human tends to be able to recognize buildings solely by their shapes. Different shape features have been exploited as a powerful clue for different visual object identification tasks. For example, edge direction histogram has been employed for city/landscape classification (Vailaya *et al.*, 1998) since city images

typically contain horizontal and vertical edges. In (Fritz *et al.*, 2005), the Scale Invariant Feature Transform (SIFT) descriptors (David, 2004) combined with feature selection mechanism have been proposed to detect buildings. Recently, a local shape descriptor called Histogram of Oriented Gradient (HOG) was proposed (Dalal and Triggs, 2005) by using the spatial distribution of edges and the applications to a number of different object detections such as human detection have been reported with excellent results. For any given subwindow of the image, a HOG accumulates over each edge point, the angles of the gradients that fall into a particular range. The contribution of an edge point to the HOG is weighted by the gradient magnitude at that point.

How to create a CBIR system for historic Chinese architecture images is an important area of research bridging society, culture and information technology. Ancient Chinese architecture is an important component of the system of world architecture. During its long development, it gradually formed its characteristics which featured timberwork combining stone carving, rammed earth construction, bucket arch buildings and many other techniques. Industrious Chinese laboring people created many architectural miracles such as the Great Wall, Forbidden City and the Mausoleum of the First Qin Emperor. There were many different styles of ancient Chinese architecture such as Buddhist architecture, Islam mosques, Taoist architecture, Chinese temples, imperial architecture, garden architecture and general architecture. All of them are unique and equally exquisite. Querying by example the databases of Chinese architecture of historical value would facilitate both educational and research and enable the exploration of unknown inter and intra cultural associations.

For architectural image retrieval, it is also important to represent the spatial relations between local shapes. In a CBIR system for ancient Chinese architectures, a strong match of two images goes beyond the distribution of edges and also involves a spatial correspondence. To this end, a pyramid extension of HOG called PHOG which impose grids of various resolutions on the image space can offer more efficient image matching (Bosch *et al.*, 2007).

While HOG counts occurrences of gradient orientation in localized portions of an image, PHOG captures perceptually salient features and suggest that locally orderless matching may be a powerful mechanism for estimating overall perceptual similarity between images. It has been proved experimentally that representations of shape by PHOG often perform as well as or even better than local appearance patches.

Though, content-based image retrieval has made enormous progress as the size of image databases grows dramatically, how to manage the use of massive information is still very acute. Most of existed CBIR systems calculate the similarity between the query image and all the images in the database and rank the images by sorting their similarities, thus cost significant query processing time when the number of images in the database is large. To improve the efficiency, image classification is a useful technique that can exploit a priori information regarding the organization of the images. For example, it can help associates each image in database with a class label such that the images associated with the same label are similar to each other.

On the other hand, classification can capture high level semantic concepts by learning the way that images of the same semantics are similar. In this study a further investigation concerning acceleration of retrieving by grouping similar images into classes is made by applying the Support Vector Machine (SVM) (Cristianini and Shawe-Taylor, 2000) which has shown promise in a variety of classification tasks. Based on a variation of regularization techniques for regression, SVM seeks a globally optimized solution and avoids over-fitting thus showing the ability to construct predictive models with larger generalization power.

SVM is originally designed for binary classification and the extension of SVM to the multi-class scenario is still an ongoing research topic (Hsu and Lin, 2002). The conventional way is to decompose an M-class problem into a series of two-class problems and construct several binary classifiers. The earliest and one of the most widely used implementations is the one-against-all method which constructs M SVM classifiers with the i-th one separating class i from all the remaining classes.

MATERIALS AND METHODS

Feature extraction by pyramid histogram of oriented gradients: Many types of shape and perceptual grouping features have been proposed to compute statistics of the primitives present in images. A state of the art gradient descriptor Histograms of Oriented Gradients (HOG) (Fig. 1) (Dalal and Triggs, 2005) is based on evaluating a dense grid of well-normalised local histograms of image gradient orientations over the image windows. Specifically, an image is first divided into non-overlapping pixel regions, or cells. For each cell a 1D histogram of gradient orientations over pixels in that cell is accumulated. The gradient at each pixel is discretized into one of nine orientation bins and each pixel votes for the orientation of its gradient with a strength that depends

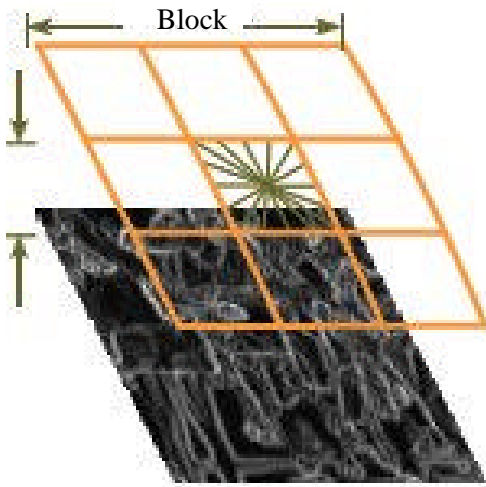


Fig. 1: A rectangular HOG descriptor with 3×3 blocks of cells

on the gradient magnitude. Finally, the histogram of each cell is normalized with respect to the gradient energy in a neighborhood around it.

The HOG feature only encodes the gradient orientation of one image patch without consideration of where this orientation is from in this patch. Therefore, it is not discriminative enough when there is crucial spatial property of the underlying structure of the image patch. The objective of the Pyramid Histogram of Oriented Gradient (PHOG) (Bosch *et al.*, 2007) is to take the spatial property of the local shape into account while representing an image by HOG. The spatial information is represented by tiling the image into regions at multiple resolutions based on spatial pyramid matching (Lazebnik *et al.*, 2006). The idea is shown in Fig. 2. Each image is divided into a sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions in each axis direction. The number of points in each grid cell is then recorded. The number of points in a cell at one level is simply the sum over those contained in the four cells it is divided into at the next level thus forming a pyramid representation (Bosch *et al.*, 2007). The cell counts at each level of resolution are the bin counts for the histogram representing that level. The soft correspondence between the two point sets can then be computed as a weighted sum over the histogram intersections at each level.

For each grid cell at each pyramid resolution level, a HOG vector is computed. The final PHOG descriptor for the image is then a concatenation of all the HOG vectors. In the implementation we follow the practice in (Bosch *et al.*, 2007) by limiting the number of levels to

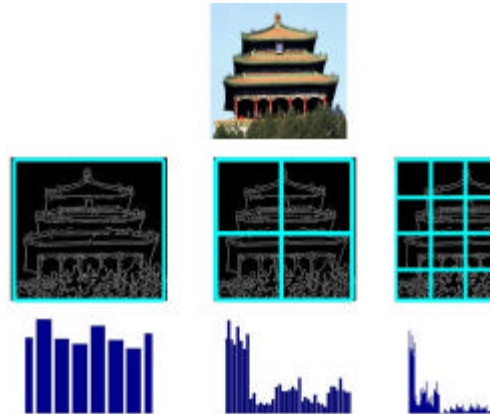


Fig. 2: A schematic illustration of PHOG. The descriptor consists of a histogram of orientation gradients over each image subregion at each resolution level

$L = 3$ to prevent over fitting. The PHOG is normalized to sum to unity. This normalization ensures that images with more edges for example those that are texture rich or are larger are not weighted more strongly than others.

PHOG is not the same as a scale space pyramid representation of edges (Lazebnik *et al.*, 2006) as there is no smoothing between levels of the pyramid and all edges are computed on the high resolution image. The PHOG descriptors can be compared using the same histogram intersection method as the pyramid match or by other means such as the χ^2 distance between the histograms at each level.

Image categorization by multiclass SVM: The performance of image retrieval can be much improved if images in the database could be classified into several categories first. A typical CBIR system is mainly based on the ranking of feature similarities between the query image and any target images which has the important problem of semantic gap in the sense that the relatively limited descriptive power of low level imagery features does not always agree with the richness of user semantics. By grouping images into meaningful categories by introducing interpretations for high feature similarities among images, the semantic gap can be substantially reduced.

Image classification can also enhance the retrieval speed, since a large image database can be organized according to a priori information regarding the meanings of the images in the database before a query is posed. Then image search can be performed within relevant classes which certainly saves significant query processing time without compromising the retrieval

precision. Without the partition of images into different categories, the retrieval would be conducted based on a whole database comparison and the time of image retrieval will depend in a large degree on the number of images in the database.

It have collected >8000 images of some famous Chinese historical architectures, taken from places such as the Forbidden City, city of Qufu (the birthplace of Confucius), the Great Wall, mausoleums, Pingyao old town (in Shanxi Province) and many other historical spots. Many types of Buddhist architecture such as temple and pagoda have been included. Images of some well-known old bridges, particularly the Zhaozhou Bridge and Lugou Bridge were added into the collections. Some special Chinese ancient architectures have also been taken into account. Examples include (Hiremath and Pujari, 2007). Paifang or Pailou which is traditional Chinese architectural form like an archway. Originally served as a marker for the entrance of a building complex or a town, it had evolved into a purely decorative monument (Bowyer and Flynn, 2000). Huabiao, a marble ornamental pillar engraved with entwisting dragons and auspicious clouds used to decorate important buildings in old China (Liu *et al.*, 2007) gloriette, pavilion in a garden from which views may be enjoyed. There are three components that make up the foundation of ancient Chinese architecture, the foundation platform, the timber frame and the decorative roof. Therefore, some close-up images from historical architectures was also gathered particularly the pictures of stele, roof, eaves, queti, girder, corridor and hall.

After collecting over 1000 ancient Chinese architecture images, all the images categorized into 20 classes for example: palace, temple, hall, the Temple of Heaven, stele, rampart, roof, eaves, belfry, queti, Paifang, bridge, pagoda, gloriette, garden, caisson, corridor, girder. Some of the examples are shown in Fig. 3.

Many different classification schemes have been studied in the past for image classifications. For example, texture image retrieval and face image retrieval have been investigated together with their respective classification issues. Examples of image classifiers include the k-nearest neighbor, the decision tree, the Bayesian net, the maximum likelihood analysis, the linear discriminant analysis, the neural network, etc.

Support vector machine: a short introduction: The basic idea of support vector machine is to determine a classifier (or regression) machine in a high dimensional feature space by mapping the input space to a high dimensional feature space. Some so called support vectors are found that represent the training data and then will be formed into a model by an SVM algorithm. For binary pattern

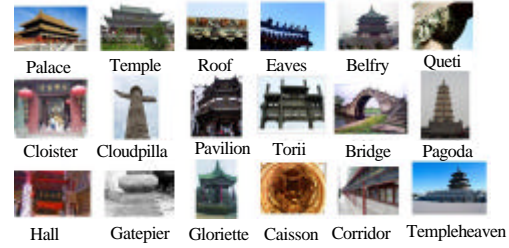


Fig. 3: Some representative ancient Chinese architecture images

classification, the essence of SVM is to find the optimal separating hyperplane that separates the positive and negative examples with maximal margin. Usually, the classification decision function for a linearly separable problem can be represented by:

$$f = \text{sign}(w \cdot x + b) \quad (1)$$

which determines the classification decision function by minimizing the empirical risk:

$$R = \frac{1}{l} \sum_{i=1}^l |f(x_i) - y_i| \quad (2)$$

where, l represent the size of examples.

SVM is based on the structural risk minimization principle (Furey *et al.*, 2000). The optimal separating hyperplane is determined by giving the largest margin of separation between different classes.

This optimal hyperplane bisects the shortest line between the convex hulls of the two classes. The optimal hyperplane is required to satisfy the following constrained minimization as:

$$\begin{aligned} &\text{Minimize: } \frac{1}{2} w^T w \\ &\text{subject to: } y_i (w \cdot x_i + b) \geq 1, i = 1, \dots, l \end{aligned} \quad (3)$$

The SVM classifier is then obtained by the inner product $x_i^T x$ where $i \in S$ the set of support vectors. However, it is not necessary to use the explicit input data to form the classifier. Instead, all that is needed is to use this inner products between the support vectors and vectors of the feature space. That is, by defining the kernel:

$$K(x_i, x) = x_i^T x \quad (4)$$

And a nonlinear classifier can be obtained as:

$$f(x) = \text{sign}\{\alpha_0 y_i k(x_i, x) + b_0\} \quad (5)$$

There are three most popular kernels, the polynomial, Gaussian and the tanh kernel. The polynomial kernel of degree d is given by:

$$k(x, z) = (x \cdot z + c)^d \quad (6)$$

where, c is a constant, d can be user defined. When d is 1, the kernel becomes linear. The Gaussian kernel is:

$$k(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (7)$$

Where the parameter controls the support region of the kernel. The tanh kernel is given by:

$$k(x, z) = \tanh(z \cdot x + b) \quad (8)$$

The aforementioned support vector machines were primarily designed for binary pattern classification problems. Multi-class (k -class, $k > 2$) classification can be obtained by the combination of binary classification. There is a relationship between binary classification and multi-classification and a variety of schemes have been proposed in the literature for solving multi-class problem (Franc and Hlavac, 2002).

Some of them are: one-against-one, one-against-all and directed acyclic graph SVM (DAGSVM). In the one-against-all approach, an SVM is constructed for each class by discriminating that class against the remaining ($k-1$) classes.

The number of SVMs used in this approach is k . A test pattern x is classified by using the winner-takes-all decision strategy i.e., the class with the maximum value of the discriminant function $f(x)$ is assigned to it.

In the one-against-one strategy, one SVM is constructed for each pair of classes. Thus, for a problem with k classes, $k(k-1)/2$ SVMs are trained to distinguish the samples of one class from the samples of another class. Usually, classification of an unknown pattern is done according to the maximum voting where each SVM votes for one class.

In literature Hsu and Lin (2002) gave a comparison of these methods. For most applications, the one-against-one approach is more computationally intensive since the results of more SVM pairs ought to be computed. Due to this consideration, one-against-all approach was applied.

Some conventional multi-class pattern classification methods compared: In machine learning, there are some

conventional methods that have been extensively applied for multi-class classification problems (Duda *et al.*, 2001) for example, the k -nearest neighbor classifier (kNN), the Nearest Mean Classifier (NMC), Fisher Linear Discriminant and artificial neural networks (e.g., multilayer Perceptrons). In the following we briefly summarize a few of the most commonly used methods for comparison purpose.

k-nearest neighbor classifier: k -Nearest Neighbor (kNN) classifier is a prototype-based classifier among the oldest types of classification methods. It is based on a distance function for example, the Euclidean distance for pairs of data samples. The kNN classifies a test sample on the basis of the training set by first finding the k closest samples in the training set and then predicting the class by majority voting. In other words, the class that is most common among those k neighbors is chosen as the predicted label. Obviously the kNN classifier needs to access all learning data at the time when a new test case is to be classified.

Nearest mean classifier: kNN which uses all training data to label a test sample, the NMC abstract training data first by only storing the mean of each class i.e., one prototype per class. It then classifies a test sample with the label of the nearest class prototype.

Fisher linear discriminant: In Fisher Linear Discriminant (FLD), each class is modeled by a multivariate Gaussian where each class is assumed to have the same covariance matrix. Specifically FLD finds Gaussian distributions of the data by maximum likelihood estimations for several parameters: the prior probability of class k ; the mean of class k and the common covariance matrix. Observations can be classified to the class of the nearest mean vector according to Mahalanobis distance or by a likelihood ratio test which reduces to a particularly simple form: $x \cdot w > t$ where, t is a decision threshold and w is a parameter vector computed from the Gaussian parameters.

Discriminant analysis can be implemented by using Bayesian estimation (Duda *et al.*, 2001; Theodoridis and Koutroumbas, 2003). By assuming normal densities, a quadratic classifier between the classes of the dataset can be derived.

Multi-layer perceptron: The Multilayer Perceptrons (MLP) is a common type of neural network classifier which is often trained by the error back-propagation algorithm (Haykin, 1998). MLP usually consists of a layer

of input nodes each linked by weighted connections to every one of a layer of hidden nodes, each of which is linked in turn to a set of output nodes. It has been shown that MLPs can virtually approximate any function with any desired accuracy provided that enough hidden units and enough data are given (Haykin, 1998). Therefore, it can also implement a discrimination function that separates input data into classes.

Such an ability of an MLP to learn from data makes it a practical classifier for many classification tasks. Though, successful for many applications however, MLP classifier has several limitations and training an MLP network involves a considerable degree of empiricism. And the performance often depends on the nature and quality of the data on which it is trained. For example, the classification accuracies may be sensitive for different class frequencies in the training set.

RESULTS AND DISCUSSION

Experiments: The image database used in this study contained >8000 images from different sources. The images belonging to a certain category were perceived to be similar in terms of the architectural styles. The PHOG algorithm is applied to all the classes of collected images with four levels of pyramids and 8 bins in each level. In forming the pyramid, the grid at level 1 has 2^1 cells along each dimension. Consequently, level 0 is represented by a K-vector corresponding to the K bins of the histogram, level 1 by a 4K-vector etc. Therefore, the four level PHOG descriptor of an image is a vector with dimensionality:

$$8 \sum_{l=0}^3 4^l = 680$$

Like many applications, the number of variables in the data set needs to be reduced due to the existence of irrelevant, noisy and redundant information which has been proved to be the detrimental elements leading to the inaccuracies in classification performance. Moreover, as the number of features used for classification increases, the number of training samples required for statistical modelling and/or learning systems grows exponentially. Principal Component Analysis (PCA) is an efficient dimensionality reduction technique which carry out linear transformation of data and project it to a lower dimensional subspace in such a way that most of the information is retained while discarding the noisy component of data. Similar to the concept of PCA-SIFT

introduced (Ke and Sukthankar, 2004) to reduce the dimensionality of SIFT features, SIF features applied together with PHOG. The experiment settings for all the classifiers are summarized as follows. For MLP, it experimented with a three-layer network with the error back-propagation algorithm.

Specifically, the number of inputs is the same as the number of features (i.e., 680 for raw PHOG data or the first several PCA projections), one hidden layer with 20 units and a single linear unit representing the class label. It is our observation that varying the number of hidden units in such an MLP usually does not significantly change the performance. The support vector machine classifier were optimized by quadratic programming (Cristianini and Shawe-Taylor, 2000). For kNN classifier, we chosen $k = 3$ after testing a range of different values of k with the similar results.

There are many standard procedures to test the performance of a pattern classification scheme. The commonly used ones are k-fold cross-validation and leave one out methods. The k-fold cross-validation is an established technique for estimating the accuracy of a classifier. In general, all of the examples are partitioned into k subsamples of which the k th subsamples is retained for testing the model while the remaining $k-1$ subsamples are used as training data.

Cross-validation is then repeated k times with all of the k subsamples used exactly once as the validation data. The cross-validation estimation of accuracy is the overall number of correct classifications divided by the number of instances in the data set. The hold out method is the simplest kind of cross-validation (2-fold) with the data set being separated into training set and testing set. In the following, we mainly conducted experiments by hold out methods.

Classification performance from holdout experiment: In the first experiment we compared a SVM classifier with several other methods including kNN, NMC, QDC and MLP based on the random divisions of the dataset into a training set and a test set. The SVM applied is based on the One-Against-All (OAA) scheme with values of the regularization parameter $C = 100$ and sigma parameter ($\sigma = 1$) by using the radial basis function kernel. The values are from the so-called grid search (Hsu and Lin, 2002). About 100 random splitting were repeated and the classification results on the testing set were recorded and averaged. For each random testing, the same set of training/testing were applied to the five classifiers. The experiments were conducted with the first 200 PCA

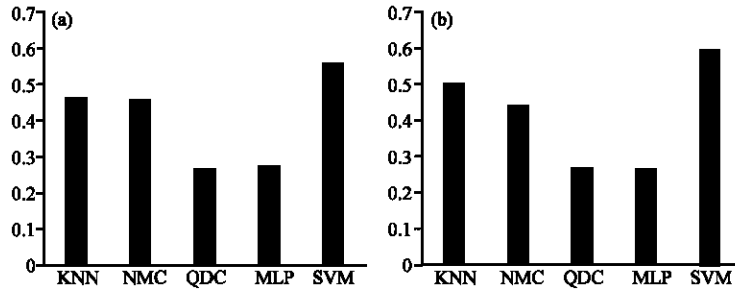


Fig. 4: Barplots of the classification accuracies from the five different classifiers. The results were from the average of 100 tests. The SVM used is based on DAG scheme. (a) 50% of data were used for training while the remaining for testing and (b) 80% of data were used for training while the remaining for testing

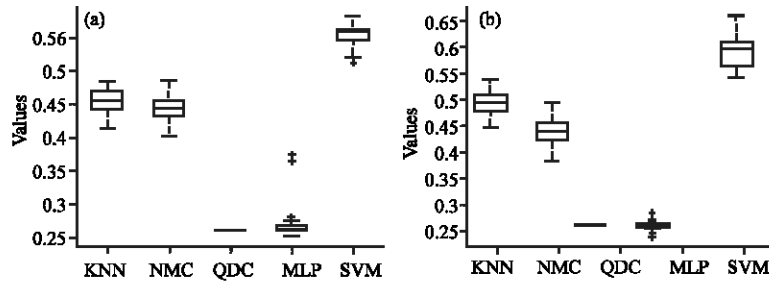


Fig. 5: Boxplots of the classification accuracies from the five different classifiers. The results were from the average of 100 tests. The SVM used is based on OAA scheme. (a) 50% of data were used for training while the remaining for testing and (b) 80% of data were used for training while the remaining for testing

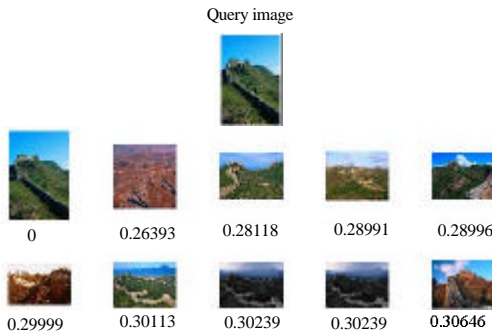


Fig. 6: A typical retrieval result from queries for great wall images



Fig. 7: A typical retrieval result from queries for images of city gate tower

scores. The results in the two subfigures (a) and (b) correspond to applying 50% for training and 50% for testing and applying 80% for training and 20% for testing, respectively (Fig. 4a, b).

It can be observed that support vector machine gives the best classification performance with regard to the accuracy (>55%) and the QDC classifier gives the worst result (<30%). The same set of experiments can also be summarized in box plots as shown in Fig. 5a, b which gives the statistical means and standard deviations of

accuracy over the 100 repeated random sampling tests. The SVM used is based on DAG scheme; (a) 50% of data were used for training while the remaining for testing and (b) 80% of data were used for training while the remaining for testing.

Retrieval results: For the sake of simplicity of evaluation, all the query images are from the database. A retrieved image is considered correct if it belongs to the same category as the query image. Each image of the 20

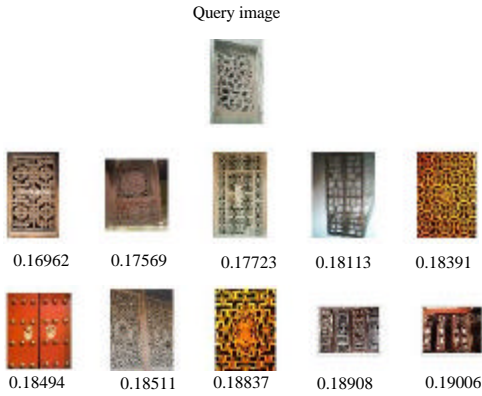


Fig. 8: A typical retrieval result from queries for images of carved door

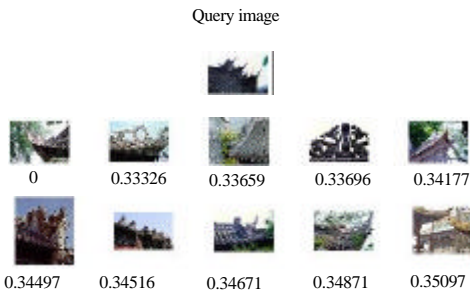


Fig. 9: A typical retrieval result from queries for images of eave

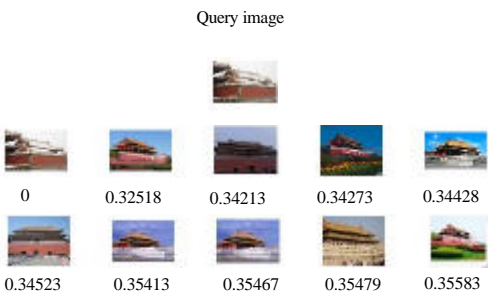


Fig. 10: A typical retrieval result from queries for images of palace

semantic categories used in this study served as the query image in turn. The following Fig. 6-11 show results from queries for great wall, city gate tower, carved door, eave, palace, ornamental columns (Hua Biao) in the 8000 image data set. The top row is the query image. The images listed in the next two rows are the top ten images returned by the query. These images are displayed based on their ranks (PHOG histogram distance between the query and image in the data set) which are similar to the



Fig. 11: A typical retrieval result from queries for images of ornamental columns (Hua Biao)

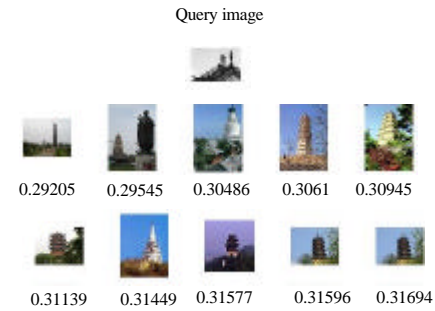


Fig. 12: An incorrect retrieval result from queries for images of roof

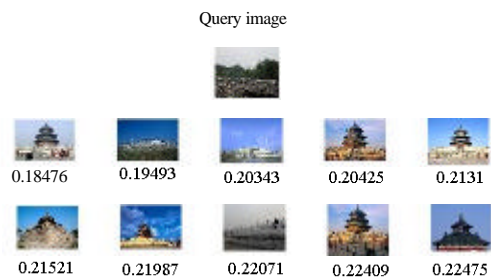


Fig. 13: An incorrect retrieval result from queries for images of old town (with mountain background)

query image in a semantic sense in that all of them belong to the same categories as the query. There are still many unsatisfactory retrieval results from the experiments. The following Fig. 12-14 show some examples which results from queries for roof, old town (with a mountain

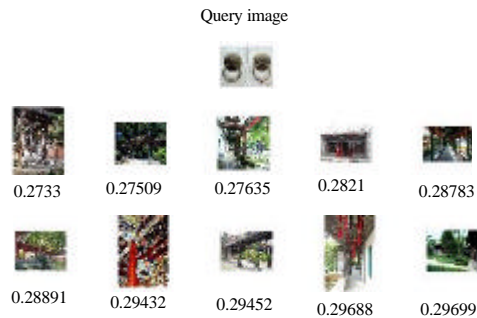


Fig. 14: An incorrect retrieval result from queries for images of door

background) and door which reflect the fact that there is always a semantic gap between a low-level image feature and a high level semantics even though PHOG is an much improved shape descriptor.

CONCLUSION

In this study, researchers propose an effective content-based image retrieval system for historical Chinese architecture image database, with three components combined to make the system efficient.

Firstly, a newly proposed local shape descriptor called Pyramid Histogram of Oriented Gradients, together with the dimension reduction scheme PCA is applied to represent shape information conveyed in an image.

Secondly, a categorization of images is pre-defined to introduce domain knowledge for the semantic description; The partitioning of the image data set into subspaces of similar elements also facilitates the restriction of search space to provide adequate response time.

Thirdly, the Support Vector Machine is applied to classify an query image followed by the ranking of the PHOG matches. Experimental results was used to show the capability of the proposed method to historical Chinese architecture image database. The integration of multiple features for better image characterization and the combination of ontology with the proposed method to reduce the semantic gap will be the future research.

ACKNOWLEDGEMENT

The research is supported by the China National Scientific and Technological Support Foundation 2006BAK31B03.

REFERENCES

- Bosch, A., A. Zisserman and X. Munoz, 2007. Representing shape with a spatial pyramid kernel. Proceedings of the 6th ACM International Conference on Image and Video Retrieval, July 09-11, Amsterdam, The Netherlands, pp: 401-408.
- Bowyer, K. and P. Flynn, 2000. A 20th anniversary survey: Introduction to content-based image retrieval at the end of the early years. *IEEE Trans. PAMI*, 22: 1348-1348.
- Cristianini, N. and J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. 1st Edn., Cambridge University Press, Cambridge.
- Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. *IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition*, 1: 886-893.
- David, G.L., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60: 91-110.
- Duda, R.O., P.E. Hart and D.G. Stork, 2001. *Pattern Classification*. 2nd Edn., John Wiley and Sons, Inc., New York.
- Franc, V. and V. Hlavac, 2002. Multi-class support vector machine. *Proc. 16th Int. Conf. Patt. Recognit.*, 2: 236-239.
- Fritz, G., C. Seifert, M. Kumar and L. Paletta, 2005. Building Detection from Mobile Imagery Using Informative SIFT Descriptors. In: *Lecture Notes in Computer Science*, kalviainen, H. (Ed.). Springer Berlin, Heidelberg, ISBN: 978-3-540-26320-3, pp: 629-638.
- Furey, T., N. Cristianini, N. Duffy and D. Bednarski, 2000. Support vector machine classification and validation of cancer tissue Samples using microarray expression data. *Bioinformatics*, 16: 906-914.
- Haykin, S., 1998. *An Introduction to Neural Networks: A Comprehensive Foundation*. 2nd Edn., Prentice-Hall, Upper Saddle River, NJ.
- Hiremath, P.S. and J. Pujari, 2007. Content based image retrieval using color, texture and shape features. *Proceeding of the 15th International Conference Advanced Computing and Communications*, Dec. 18-21, Indian Institute of Technology Guwahati, India, pp: 780-784.
- Hsu, C.W. and C.J. Lin, 2002. A comparison on methods for multi-class support vector machines. *IEEE Trans. Neural Netw.*, 13: 415-425.
- Ke, Y. and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors. *Comput. Vision Pattern Recognition*, 2: 506-513.

- Lazebnik, S., C. Schmid and J. Ponce, 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proc. IEEE Conf. CVPR, 2: 2169-2178.
- Liu, Y., D.S. Zhang, G.J. Guo and W.Y. Ma, 2007. A survey of content-based image retrieval with high-level semantics. Pattern Recogn., 40: 262-282.
- Theodoridis, S. and K. Koutroumbas, 2003. Pattern Recognition. Academic Press, Inc., San Diego, Calif.
- Vailaya, A., A. Jain and H. Zhang, 1998. On image classification: City vs. landscape. Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries, June 21, Santa Barbara, CA., pp: 3-8.
- Veltkamp, R.C., H. Burkhardt and H. Kriegel, 2001. State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer Academic Publisher, Springer, Boston, Dordrecht, London, ISBN: 1402001096, pp: 343.