

Content Based Retrieval and Classification of Cultural Relic Images[§]

Na Wei¹, M. Emre Celebi², Guo-Hua Geng¹

¹ Dept. of Computer Science, Northwest University at Xi'an,
Shanxi, 710069 China

weinaok@126.com, ghgeng@nwu.edu.cn

² Dept. of Computer Science and Engineering, University of Texas at Arlington,
Arlington, TX 76019 U.S.A.

celebi@cse.uta.edu

Abstract. In this paper we present a novel system for content-based retrieval and classification of cultural relic images. First, the images are normalized to achieve rotation, translation and scaling invariant similarity retrieval. After normalization, a combination of color and shape features is extracted from the images. In order to improve the retrieval efficiency, a modified version of principal component analysis is used to reduce the dimensionality of the feature space. Retrieval performance of the system is evaluated for three different distance functions using the normalized recall measure. A multi-class support vector machine (SVM) classifier is used for classification. The results demonstrate that the system is both effective and efficient.

1 Introduction

Content-based image retrieval (CBIR) has been a topic of considerable interest in various applications domains such as trademark retrieval, fashion design, journalism, digital museum browsing, etc. [1,2]. Numerous CBIR systems have been developed to address these problems both in the industry and academia [8,11,12].

We have developed a digital archaeology museum, which contains 2740 cultural relic images of high quality. Our project aims at advancing information technologies related to the preservation, retrieval, and dissemination of digital imagery of cultural heritages. This paper focuses on content-based retrieval and classification of cultural relic images in our digital archeology museum.

The rest of the paper is organized as follows: Section 2 presents the image preprocessing phase. Section 3 describes the feature computation, dimensionality reduction, and retrieval performance evaluation procedures. Section 4 describes the support vector machine classification. Finally, Section 5 concludes the paper and provides future research directions.

[§] Na Wei and Guo-Hua Geng are supported by the *Remote Resource Construction* project funded by the Ministry of Education of China.

2 Preprocessing

During data acquisition, the camera orientation or camera-to-object distance may change. Hence the image fed to the system may contain a pattern that is rotated, scaled, or translated compared to its original form. In such cases, either the system should use features that are invariant to such transformations or there should be a preprocessor to maintain the rotational, scaling, and translation invariancy [3]. This study uses the latter approach.

Figure 1 shows a sample of images from our database. First, the images are binarized using dynamic thresholding method described in [13]. Let $f(x,y)$ and $c=[C_x, C_y]^T$ denote the binary image function and the center of gravity of the object, respectively. To achieve translation invariance, we translate the object so that the center of gravity coincides with the origin:

$$f_T(x_i, y_j) = f(x_i - C_x, y_j - C_y) \quad (1)$$

The term radius for a pixel is defined to be the length of the straight line connecting the pixel and the origin. The scaling process will bring a radial boundary to the pattern in the image while adjusting the number of on-pixels. It thus alleviates any possible pattern deformation caused by scaling. The average radius is computed as:

$$\bar{r} = \frac{\sum_{i=1}^N \sum_{j=1}^N f_T(x_i, y_j) \cdot \sqrt{x_i^2 + y_j^2}}{\sum_{i=1}^N \sum_{j=1}^N f_T(x_i, y_j)} \quad (2)$$

In order to achieve scale invariance we scale the object with a scaling factor $s = \bar{r}/R$, where R is equal to one-fourth of the grid size:

$$f_{TS}(x_i, y_j) = f_T(s \cdot x_i, s \cdot y_j) \quad (3)$$

To achieve rotational invariance we determine the major axis orientation. Thus, the rotation angle in terms of the central moments of the object is given by:

$$\theta = \tan^{-1} \left(\frac{\mu_{20} - \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2\mu_{11}} \right) \quad (4)$$



Figure 1. Sample cultural relic images from four different classes

3 Content Based Retrieval

3.1 Feature Computation and Dimensionality Reduction

For each image in the database, we compute the following 19 features: Circularity [4], bending energy [4], 7 moment invariants $\{\Phi_1, \Phi_2, \dots, \Phi_7\}$ [4], contour sequence moments $\{F_1, F_2, F_3, F_4\}$ [4], and 1st & 2nd order color moments [5] computed from individual channels of the HSV color space. Since these descriptors are widely used in the literature, we omit their formulae.

As a result of the feature computation step, we have feature space with 19 dimensions. Well known problems associated with high dimensionality include (a) high computational cost, (b) classifier accuracy degradation, and (c) difficulty in visualization [10]. In this study we use a modified version of the Principal Component Analysis (PCA) for dimensionality reduction. In the classical PCA method, the data is represented by N principal components (PCs), where $N < K$ = the number of features. However, we still need all of the K original features to compute the N PCs since each PC is a linear combination of the original features. This modified version of PCA differs from the classical one in that instead of using N PCs , we select N original features to represent the data. We use Jolliffe's B4 Method [6] for selecting N representative features from K original features. In this method, starting from the largest PC , we select the feature with the highest coefficient on the relevant PC to represent that component, unless it has been chosen to represent a larger PC . In this way, a total of N features are retained.

The eigenvalue and the percentage of explained variance for each PC are given in Table 1.

Table 1. Eigenvalue and percentage of explained variance for each $PC/Eigenvector$

<i>PC/Eigenvector</i>	<i>Eigenvalue</i>	<i>Variance%</i>	<i>Cumulative Var.%</i>
1	7.5273	39.617	39.617
2	3.7591	19.785	59.402
3	2.0371	10.722	70.124
4	1.6328	8.594	78.718
5	1.3688	7.204	85.922
6	0.8469	4.457	90.379
7	0.5656	2.977	93.356
8	0.3837	2.020	95.376
9	0.2911	1.532	96.908

In our analysis we decided to keep 9 PCs that account for 96.9% of the variation in the original data. Using Jolliffe's B4 method we decide to keep the following original

features: Circularity, F_4 , Φ_1 , Φ_5 , Φ_7 , the 1st order color moments for H and V , and the 2nd order color moments for H and S . The remaining 10 features are discarded.

3.2 Retrieval Performance Evaluation

An important step between feature computation and similarity measurement is feature normalization. In this study, we use z-score normalization.

Retrieval performance of the system is evaluated using the normalized recall (NR) measure [8]. NR is formulated as $AVRR/IAVRR$, where $AVRR$ is the average rank of all relevant, displayed images and $IAVRR$ is the ideal average rank which is the maximum when all relevant images are retrieved on the top. NR is equal to 1 for perfect retrieval and is greater than 1 for all other cases. For the experiments we selected 98 query images from 5 different classes. The relevant images in the database for each query image are determined manually. A list of 32 best matches is displayed for each query. Figure 2 (a) and (b) show snapshots of the query interface before and after query execution, respectively.

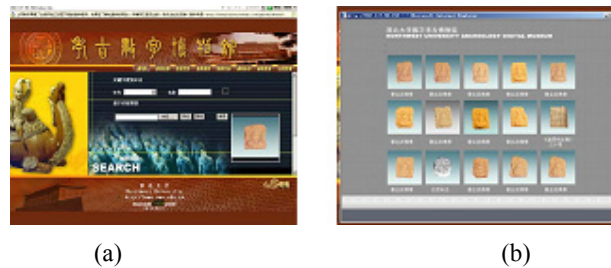


Figure 2. Snapshots of the query interface (a) before query execution (b) after query execution

The choice of distance function can drastically influence the retrieval performance [7]. For this reason, we compare the performance of three different distance functions: Euclidean, chi-square, and $K-L$ divergence. Figure 3 shows the NR values for each combination of distance function and query image class. It can be seen that Euclidean distance function outperforms both chi-square and $K-L$ divergence distance functions.

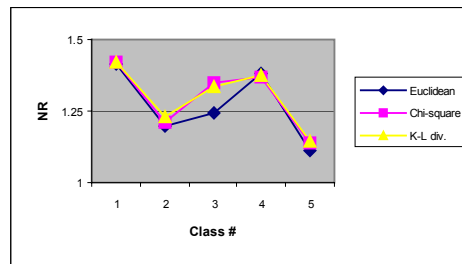


Figure 3. Retrieval performance comparison using the NR measure

4 Support Vector Machine Classification

Support vector machines (SVMs) are non-parametric supervised classifiers with excellent generalization capabilities. They have been successfully used in numerous applications such as hand-written digit recognition, object recognition, speaker identification, face detection, etc. [9].

SVMs are originally designed for binary classification. There are two main approaches for extending the basic SVM to handle multi-class classification: (1) Modifying the design of the SVM in order to incorporate the multi-class learning directly in the quadratic solving algorithm (2) Combining several binary classifiers: “One against one” applies pairwise comparisons between classes, while “One against the others” compares a given class with all the others put together. In this work we experimented with the second approach. In our experiments, the “one against the others” approach often produced quite unbalanced two-class problems. Therefore, we chose the “one against one” method to perform multi-class classification.

In a k -class classification problem, the “one against one” method requires $k(k-1)/2$ binary classifiers, one for each pair of classes. In the training phase, each binary classifier is trained to classify two classes. In the testing phase, the output from each binary classifier in the form of a class label is obtained. The class label that occurs the most is assigned to the data.

The general purpose Gaussian RBF is chosen as the kernel:

$$K(x, y) = \exp(-\gamma(x - y)^2) \quad (5)$$

For $k = 2$ and $k = 5$, i.e. 2-class and 5-class problems, we set the kernel parameter γ as 0.5 and 0.2, respectively and estimate the classification error for each cost parameter (C) value in the set $\{2^{11}, 2^{10}, \dots, 2^{-2}\}$. For each (C, γ) combination, the validation performance is calculated using 70% of the data as the training set and 30% as the test set. Then, we calculate the average classification error rate using the leave-one-out method with the (C, γ) combination that achieves the highest validation rate. Table 2 shows the optimal (C, γ) value and the corresponding average classification error rate for the 2-class and 5-class cases.

Table 2. Average classification error rates

<i>Problem</i>	(C, γ)	<i>Error rate</i>
2-class	$(2^4, 0.5)$	2 %
5-class	$(2^7, 0.2)$	4%

5 Conclusions

In this paper we presented a novel system for content-based retrieval and classification of cultural relic images. First, the images are normalized to achieve rotation, translation, and scaling invariant similarity retrieval. Following image normalization, a combination of color and shape descriptors is extracted from the images. A modified version of principal component analysis is employed to reduce the dimensionality of the feature space. To test the retrieval performance, three distance functions are evaluated. Finally, a multi-class support vector machine is used for classification.

Future work will be directed towards incorporation of texture features and relevance feedback to the system in order to improve the retrieval performance.

References

1. Zhou, M.Q., Wei, N., Geng, G.H.: Content Based Image and Model Retrieval Techniques. The Third Asian Workshop on Foundations of Software. (2004), Xi'an, China
2. Wei, N., Geng, G.H., Zhou, M.Q.: 3D Model Retrieval Using Distance Histogram. Proc. of the 14th National Conference on Artificial Neural Networks and Natural Computing. (2004), HeFei, China. 537-539
3. Yuceer, C., Oflazer, K.: A Rotation, Scaling, And Translation Invariant Pattern Classification System. Pattern Recognition. 26 (1993) 687-710
4. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis, and Machine Vision. Brooks/Cole Pub. CA (1998)
5. Stricker, M., Dimai, A.: Color Indexing with Weak Spatial Constraints. Proc. of the SPIE. 2670 (1996) 29-40
6. Dunteman, G.H.: Principal Component Analysis. Sage Publications. (1989)
7. Zhang D.S., Lu, G.: Evaluation of Similarity Measurement For Image Retrieval. Proc. of IEEE ICNNSP'03. (2003), Nanjing, China. 928-931
8. Faloutsos, C., Equitz, W., Flickner, M. et al.: Efficient and Effective Querying by Image Content. Journal of Intelligent Information Systems. 3 (1994) 231-262
9. Chapelle, O., Haffner, P., Vapnik, V.: Support Vector Machines for Histogram-Based Image Classification. IEEE Trans. on Neural Networks. 10 (1999) 1055-1064
10. Celebi, M.E., Aslandogan, Y.A.: Content-Based Image Retrieval Incorporating Models of Human Perception. Proc. of IEEE ITCC Conf. (2004), Las Vegas, NV. 241-245
11. Pentland, A., Picard, R.W., Sclaroff, S.: Photobook: Content-based manipulation of image databases. International Journal of Computer Vision. 18 (1996) 233-254
12. Smith, J.R., Chang, S.-F.: VisualSEEK: A Fully Automated Content Based Image Query System. Proc. of the ACM Multimedia Conf. (1996), Boston, MA. 87-98
13. Bernsen, J.: Dynamic Thresholding of Grey Level Images. Proc. of the 8th Int. Conf. on Pattern Recognition. (1986), Paris, France. 1251-1255