Content-Based Retrieval of Cultural Relic Images in Digital Museum*

Tongwei Ren[1,2] and Gangshan Wu[1,2]

[1] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093
[2] Department of Computer Science and Technology, Nanjing University, Nanjing, 210093
rentw@graphics.nju.edu.cn

Abstract. With the popularization of digital museum, effective retrieval in huge image databases of special domain has attracted much research attention. As an effect approach, corresponding semantic information is adopted in many retrieval systems. However, most cultural relic retrieval systems only supply the explicit use of semantic information. It requires the user to be professional in this domain. In this paper, we propose a novel relevance feedback method which combines semantic annotation to visual feature implicitly. With this method, the user can unknowingly use professional semantic information to retrieve images. We also do some improvement in feature extraction and similarity measurement methods to fit the retrieval basis of unprofessional user better. The experiment results show that our approach is effective and efficient.

Keywords: Content-Based Image Retrieval, cultural relic image, relevance feedback, implicit combination.

1 Introduction

With the popularization, digital museum becomes more and more important in knowledge acquiring for commonalty. Effective retrieval in the huge image databases has been an active research area in the past few years. Differing from general image sources, the image databases in digital museum focus on special domains, and the corresponding semantic information plays an important role in retrieval. So the combination of visual feature and semantic information is accepted as an effective approach. In cultural relic image retrieval, several approaches have been proposed.

Li et al. presented an approach for Dunhuang fresco in [1], which selected color, element shape and layout as the visual features and supplied semantic-based retrieval by keywords. Wei et al. presented a retrieval approach for cultural relic image database in [2], which extracted 19 features from preprocessed images and supplied retrieval based on example image and keywords. However, the above approaches only make the explicit use of the semantic information, which may be invalid when the user does not have enough professional knowledge in this domain.

* Supported by the National Natural Science Foundation of China under Grant No.60533080.

Z. Pan et al. (Eds.): ICAT 2006, LNCS 4282, pp. 1108 – 1117, 2006.

To conquer the weakness, we proposed a relevance feedback method which combines the semantic annotation to visual feature implicitly. In this method, semantic annotation is not presented but automatically analyzed based on feedback during the whole retrieval procedure. Considering visual feeling is the most important basis in requirement description and feedback operation for unprofessional user, we analyze the visual related characters of cultural relic images in digital museum, and do some improvements in feature extraction and similarity measurement to fit human-vision characters better. Finally, we implement a prototype system, and evaluate it according to the analysis of the retrieval model classification in digital museum. The experiment results show our system is efficient in all application conditions.

The rest of paper is organized as follows. Section 2 analyzes the characters of cultural relic images and the retrieval models in digital museum. Section 3 describes the preprocessing, feature extraction and similarity measurement approaches. Section 4 proposes a novel relevance feedback method combining semantic annotation to visual feature implicitly. Section 5 evaluates the retrieval approach, and presents the experiment results. Finally, Section 6 concludes the paper and provides future research directions.

2 Analysis

For visual feeling usually plays the most important role in requirement description and feedback operation for unprofessional user, we analyze the visual-related characters of cultural relic images in digital museum. Based on the analysis, we improve the feature extraction and similarity measurement methods to fit human vision characters better. We also analysis the classification of retrieval models in digital museum and design our experiment based on it.

2.1 Visual-Related Image Characters

For the characters of cultural relic and strict capture criterion, the cultural relic images in digital museums are normative and have the following visual-related characters:

- ♦ Centered object region and simplex background. To emphasize the exhibit, the exhibit usually lies near the center position and the background has simplex visual characters. The large proportion of background provides the requirement to eliminate background before feature extraction.
- ♦ Transform relations between object regions. To display the exhibit fully, several images are usually captured for the same exhibit from different directions or positions. The object regions in these images have translation, rotation, scaling or reflection relations. The relations require the visual feature descriptors to be invariable in the above aspects.
- ♦ Complex visual characters of object region. Various properties of cultural relic bring the great varieties in color, texture or shape between different exhibits. The complexities in visual characters require the features descriptors should be particular and adaptive to human vision characters.

Based on the analysis, the visual features should be extracted from object region and the descriptors should fit human vision characters and have good invariance.

2.2 Retrieval Models for Digital Museum

Base on the conclusion in [3], we classify the retrieval models in digital museum into three classes:

- ♦ Certain exhibit searching. The aim may be a certain cultural relic in user impression. Just the images of the exhibit fit the user's requirements.
- ♦ Category retrieval. The aim is searching for a special category of cultural relics which are close in appearance, material or function. And each exhibit in this category belongs to the result set.
- ♦ Browsing. The user starts his search without specific aim and modifies the aim by operating on result set.

In actual retrievals of each model, the unprofessional users usually describe their requirements by approximate image or sketch than professional keywords, and the visual feeling is always emphasized more than professional characters in object characters' description.

3 Content-Based Retrieval

Based on the analysis in 2.1, we preprocess the images by a background-forecast method and extracted visual features from object region by the improved descriptors. Then we normalize the feature vectors by Guassian model and measure the similarities by Quadratic distance and Euclidean distance.

3.1 Preprocessing

The purpose of preprocessing is eliminating the infection of background to feature extraction. Based on the background and layout characters of cultural relic images in digital museum, we propose a background-forecast method as follows:

Input: original gray image
Output: segmented matrix, in which the open-pixel set denotes the object region and the close-pixel set denotes the background region
Procedure:

1. Divide the original image into 8×8 blocks.
2. Assume the blocks on board as "background block", and compute the mean gray level of each background block as "block value".
3. Assume threshold d_G, select a block with 3 neighboring background blocks, and compute $\overline{bValue}$ as the mean value of their block values. Select the pixels whose gray level g satisfies $|g - \overline{bValue}| < d_G$, and assign the mean value of these pixels' gray levels to block value of this block. If no such pixels, assign $\overline{bValue}$ to the block value.
4. Repeat step 3 to compute the block value of each block.

5. Create the forecasted background in the same size of original image, and the gray level of each pixel equals the block value of the block which it lies in. Smooth the forecasted background by linear smooth filter.
6. Do minus operation on the original image by the forecasted background, and smooth the result by linear smoother filter.
7. Binarize the smoothed result, and get the segmented matrix. Finish.

The experiments show that the obvious difference between background and object is not always represented in gray level. It may be represented in any component of HSV model. So the above algorithm is carried out on each component of HSV model, and the best segmented matrix is automatically selected by evaluating the following aspects:

- ♦ The inner-aggregation property of object region and background region. It is measured by the standard deviations of object region and background region.
- ♦ The difference between object region and background region. It is measured by the difference between the mean values of object region and background region.
- ♦ The area proportion and location of object region. The area proportion is computed and measured whether larger than a pre-assumed threshold p_{area}. And location is measured by the distance between the centroid position of object region and the center position of image.

The proposed method can deal with the cultural relic with holes and get a good effect in experiment. Fig. 1 shows a sample of preprocessing consequence.

Fig. 1. Preprocessing Consequence

3.2 Feature Extraction

In this paper, we describe the visual feature of cultural relics by color, texture and shape. Based on the analysis in 0, we select the descriptors with good invariance and improve them to fit human-vision characters better.

In order to describe color feature, we select HSV model as color space and quantize it based on human-vision characters. We divide the color space into four parts as black, white, gray and bright color [4], and quantize the gray part and bright color part [5] further. In this way, the color space is quantized uniformly into 21 levels. Then we select color histogram as the primary descriptor and dominant color as the assistant descriptor of color feature. The experiment results show the combined color feature can emphasize the major factors in color comparison by human.

We select Co-occurrence Matrix to represent the texture feature of object region, and test the 14 descriptors [6] on rotation invariance, scaling invariance, clustering ability and partition ability. We carry out the test on 1536 images of 24 classes, which belong to Brodatz Album image library. The result show that F_{con}, F_{cor}, F_{ssv}, F_{sv}, F_{se}, F_{dv} and F_{de} have good performance in the above four aspects. We prove that F_{cor} and F_{ssv} are identical. So we compute four co-occurrence matrixes of each image on direction (0,1), (-1,1), (1,0) and (-1,-1), and calculate the rest six descriptors of each matrix. Then we compute the mean value and the standard deviation of the four values by each descriptor. Finally, we use the 12 values as the texture feature descriptor.

We select Hu Invariant Moment to represent the shape feature. We test the variances of translation, rotation, scaling and reflection of Hu Invariant Moment on the above image library. The results show ϕ_1 to ϕ_6 have good performance in the above four aspects, and ϕ_7 is invariable in the previous three aspects but not invariable in reflection. We prove the shortage of ϕ_7 in reflection is caused by its definition, and use the absolute value of ϕ_7 instead of it. Finally, we use the values of ϕ_1 to ϕ_7 as the shape feature descriptor.

3.3 Similarity Measurement

According to the relativities between the components, we use different similarity measurement methods.

To color feature, it is obvious that the relativities among each color are not the same, i.e. red is more similar to orange than purple. So we use Quadratic distance to measure the similarity of colors:

$$d^2 = (c_i - c_j)\, A\, (c_i - c_j)^T \quad . \tag{1}$$

where $A(i, j)$ denotes the similarity between color *i* and color *j*. In order to compute the color similarity matrix, we divide the color space into bright colors and un-bright colors (black, gray and white). To bright colors, we only consider the hue component, and compute the similarity between color *i* and color *j* as follows:

$$A(i,j) = \frac{\int_{iStart}^{iEnd} \int_{jStart}^{jEnd} \Delta\phi(x,y)\, dxdy}{\int_{iStart}^{iEnd} x\, dx \int_{jStart}^{jEnd} y\, dy} \quad . \tag{2}$$

where x is an arbitrary legal hue value of color *i* and y is the same to color *j*, and $\Delta\phi(x, y)$ is defined as follows:

$$\Delta\phi(x,y) = \begin{cases} 2\times|x-y|, & |x-y| \le 0.5 \\ 2\times(1-|x-y|), & |x-y| > 0.5 \end{cases} \quad . \tag{3}$$

We compute the similarity between un-bright colors on their value component, and assume the similarity between any bright color and un-bright color is zero.

We compute the dominant color distance d' using the same expression, and modify the previous distance in proportion to the value of d'. If the modified value overflows, we intercept it and let it in the range [0,1].

It is consented that the components of texture feature vector extracting from co-occurrence matrix are independent. So we use Euclidean distance to measure texture similarity. However, the experiment results show that the magnitudes of each component are obvious different. In order to exactly control the contribution to similarity measurement of each component, we use Guassian model [7] to normalize each component as follows:

$$t_i' = \frac{t_i - m}{6\sigma} + 0.5 \ . \tag{4}$$

where m denotes the mean value of the values of the component, and σ denotes the standard deviation. After the processing, 99% of texture vector component values will fall in the range [0,1], and the rest will be intercepted. In order to get available values of m and σ for changing retrieval circumstance, we make a training set with enough size and complexity, and compute the distances on each component of any image couple. Based on this, we compute m and σ values of each component.

For the component of shape feature vector extracting by Hu invariant moment is also consented independent, we use the same method to process shape feature vector and measure the similarity.

For the numbers of vector components of each feature are not the same, color, texture and shape similarities make different contributions in the integrated similarity measurement. We solve the problem by the above method. After the processing, each feature makes the same contribution to similarity measurement in default condition. In actual retrieval, the user can modify the priority of each feature to satisfy the retrieval requirements.

4 Relevance Feedback

Based on the normalization of annotation in digital museum, we extract the cultural relic property descriptions from image annotation automatically. Then we combine extracted semantic information to visual feature in relevance feedback implicitly and use clustering method to refine user requirements iteratively.

4.1 Visual Feature Clustering

To simplify the problem, we assume the user choose all related images. So the chosen images compose the positive example set, and the rest compose the negative example set. We use a 63×10 vector to present the modified requirement vector (the dimension of original feature vector is 63), which is composed of the value range and mean value of both positive example set and negative example set, positive weight, negative weight, and overlapping value region of each component. The algorithm is as follows:

Input: the visual feature vector sets of positive examples and negative examples
Output: modified requirement vector
Procedure:

1. Count the positive example number N_{pos} and the negative example number N_{neg} .
2. Select a component as the processing component, and order the value sequences on this component in positive examples and negative examples respectively.
3. To positive value sequence, process as follows:
   - ♦ Assume threshold p_{num} , select the closest $N_{pos} \times p_{num}$ values, and compute the average distance $\bar{d}_{pos}$ of the selected range.
   - ♦ Assume threshold α , for each left value, if distance between it and any boundary value of the selected range is less than $\alpha \times \bar{d}_{pos}$, add it into the range. And repeat the step till no available values left.
   - ♦ Modify the range by the value in original requirement vector and history retrieval data, and the result is the positive value range on the component.
   - ♦ Compute the mean value of positive value range m_{pos} .
4. Compute the negative value range and m_{neg} as the above steps.
5. Compute the overlapping range.
6. If the overlapping range is null, assign 1 to positive weight w_{pos} and -1 to negative weight w_{neg} . Otherwise, compute the proportions of overlapping range in positive value range and negative value range. Assume threshold p_{over} , assign 0 to the weight if the corresponding proportion is lager than p_{over} , or compute the weight according to the proportion.
7. Repeat step 2 to step 6, compute the values on each component. Finish.

Since the modified retrieval vector becomes a set of weight and value range, a new similarity measurement is required. After modification, each component affects the similarity independently. So we compute the effect of each component and compute the similarity as follows:

$$d = \sum_{i=1}^{63} e_i \ . \tag{5}$$

where e_i denotes the effect of component *i*, which is computed as follows:

- ♦ If the retrieval value falls in the positive range (not including overlapping range), assign w_{pos} to e_i .
- ♦ If the retrieval value falls in the negative range (not including overlapping range), assign w_{neg} to e_i .
- ♦ If the retrieval value falls in the overlapping range, compute e_i according to the distances from the retrieval value to m_{pos} and m_{neg} .

4.2 Implicit Combination with Semantic Annotation

Considering the annotations of culture relic images are normal in format and description, we extract some properties of exhibit as the feature. We select the properties which are described by normative and finite word sets, i.e. the material of cultural relics are described by the word set {Jade, Stone, Copper, …}. We describe such property by a vector which is composed of the candidate values, and describe any value of the property by assigning 1 to the corresponding component value (the default value of each component is 0).

The annotation feature vector is also modified during relevance feedback. Unlike visual feature vector, each component of the annotation feature vector is only two candidate values: 0 or 1. That means the value range and mean value are worthless here. So we use the positive weight w_{pos} and negative weight w_{neg} to present the modified annotation feature vector. The algorithm is as follows:

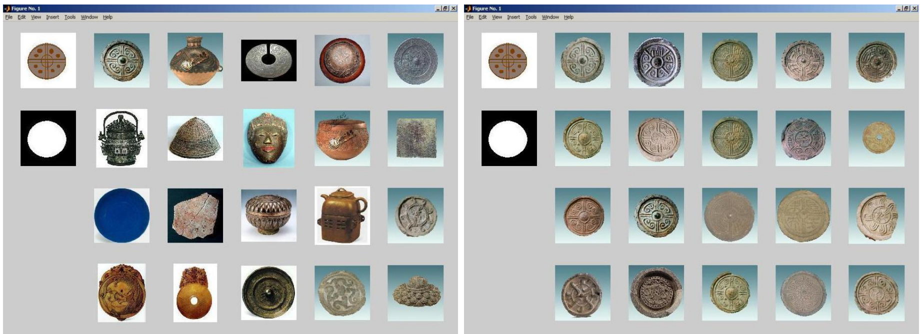
Input: the visual feature vector sets of positive examples and negative examples
Output: modified requirement vector
Procedure:

1. Assume threshold N_0, count the "1" value numbers N_i on each component in the return set.
2. If only one component is hit, let its $w_{pos}=1$ and $w_{neg}=1$. Process the rest components as follows: if $N_i<N_0$, treat the component as low- probability component, and let $w_{pos}=w_{neg}=0$; if $N_i\geq N_0$, let $w_{pos}=0$ and $w_{neg}=-1$. Finish. Otherwise, go to step 3.
3. Compute the current precision P. To each component, if $N_i<N_0$, let $w_{pos}=w_{neg}=0$; otherwise, process it as follows: compute its hit rates h_i on each component, if $h_i\geq P$, let $w_{pos}=(h_i-P)/(1-P)$ and $w_{neg}=0$; if $h_i<P$, let $w_{neg}=(h_i-P)/P$ and $w_{pos}=0$. Finish.

In similarity measurement, the effect of a component equals the nonzero one of w_{pos} and w_{neg}. If both of them are zero, then assign 0 to the effect value.

5 Experiment

We evaluate the performance of proposed approach based on the analysis in 0. We select 1680 high quality cultural relic images from the archeology digital museum of Northwest University, which contain the cultural relics of different classes. We obtain 50 nonstandard cultural relic images from internet and draw 50 sketches by hand, and combine them to be the test set. Fig. 2 shows a sample of retrieving roof tile by sketch, in which the left part shows the sketch and corresponding segmented image and the right part shows the retrieval result. Fig. 3 shows the relation between average retrieval precision and relevance feedback times. The experiment results show that our approach is effective and efficient in each retrieval model.



(a) (b)

Fig. 2. A Sample of Retrieving by Sketch. (a) shows retrieval result before relevance feedback, and (b) shows retrieval result after 3 times relevance feedback.

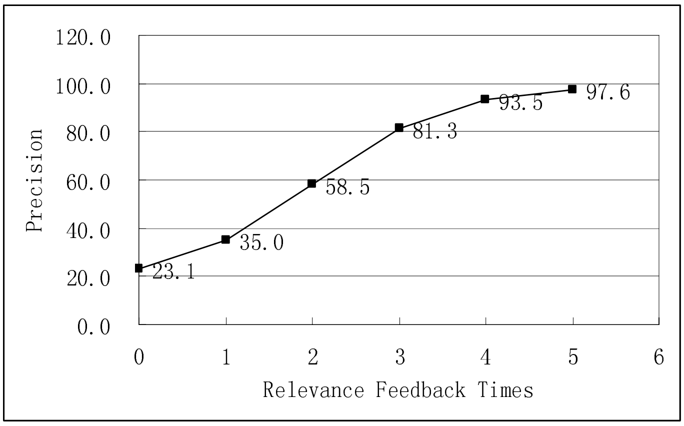


Fig. 3. Consequence Broken line diagram

6 Conclusion and Future Work

In this paper, we present a solution of making implicit use of semantic information in cultural relic image retrieval in digital museum. We propose a relevance feedback method which combines semantic annotation to visual feature implicitly. We also do some improvement in visual feature extraction and similarity measurement to fit unprofessional users' retrieval basis better. The experiment results show our approach is effective and efficient.

Our research will continue to improve the performance of relevance feedback method in part-feedback condition and to implement the auto-partition of user's retrieval model.

Acknowledgments. Our research is funded by Digital Museum of Chinese Universities and China Digital Science and Technology Museum.

References

1. X.Y. Li, D.M. Lu, Y.H. Pan, "Color Restoration and Image Retrieval for Dunhuang Fresco Preservation", Multimedia, IEEE, 7(2):38-42, 2000.
2. N. Wei, M.E. Celebi, G.H. Geng, "Content Based Retrieval and Classification of Cultural Relic Images", Proceedings of the Second International Symposium on Neural Network, Chongqing, 2005.
3. I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments", IEEE Transactions on Image Processing, 9(1):20-37, 2000.
4. D. Androutsos, K.N. Plataniotis, A.N. Venetsanopoulos, "Image Retrieval Using the Directional Detail Histogram", Proceeding of SPIE, 3312:129-137, 1997.
5. Y.M. Tian, G.S. Lin, "Retrieval Technique of Color Image Based on Color Features", Journal of XiDian University, 29(1):43-46, 2002.
6. R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, SMC3:610-621, 1973.
7. H.Y. Lee, H.K. Lee, Y.H. Ha, "Spatial Color Descriptor for Image Retrieval and Video Segmentation", IEEE Transactions on Multimedia, 5(3):358-367, 2003.