Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Technical features in the Portal to CADAL*

WU Jiang-qin (吴江琴), ZHUANG Yue-ting (庄越挺), PAN Yun-he (潘云鹤)

(*School of Computer Science, Zhejiang University, Hangzhou 310027, China*)

E-mail: wujq@cs.zju.edu.cn; yzhuang@cs.zju.edu.cn; panyh@sun.zju.edu.cn

Received Aug. 5, 2005; revision accepted Sept. 10, 2005

Abstract: China-America Digital Academic Library Project (CADAL) is a collaborative project between universities and institutes in China and the USA, which aims to provide universal access to large scale digital resources and explore the ways of applying multimedia and virtual reality technologies to digital library. The distinct characteristic of the resources in CADAL is that it not only contains one million digital books of different languages, but also contains Terabyte level multimedia resources (image, video, and so on), which are utilized for education and research purposes. So, in the Portal to CADAL, both the traditional services of browsing and searching of digital books, and the services of quickly retrieving and structurally browsing of multimedia documents should be provided. In addition, the services of visual presentation of retrieved results are required too. In this paper, the underlying novel multimedia retrieval methods as well as visualization techniques, which are used in the CADAL portal, are investigated.

Key words: CADAL, Video summarization, Multimedia retrieval, Chinese calligraphy, 3-D visualization
doi:10.1631/jzus.2005.A1249 **Document code:** A **CLC number:** TP391

INTRODUCTION

The objective of the China-America Digital Academic Library Project (CADAL) is to provide universal access to one million digital books and multimedia resources for teaching and research, and explore the ways of applying multimedia and virtual reality technologies to digital library.

Terabyte volume of multimedia data of various types of modality, such as text, image, video, animation, are available in the CADAL, which is one of the distinct characteristic of CADAL. So CADAL presents a challenge for the application of multimedia analysis and retrieval techniques. The explosive growth of digital video data is calling for more effective approaches to indexing and retrieving of video clips based on their content, so that users can browse rapidly. How to satisfy users' requirements is a very challenging problem we have to face in the CADAL project. Here we implemented a powerful multimedia analysis and summarization system. Work on content-based information retrieval (CBIR) has focused on low-level features, in the case of images, usually color, texture, and shape. The challenge for digital libraries is to develop better ways of retrieval based on the seamless integration of semantic keywords and media-specific low-level features. Recent trends include the use of relevant feedback to add information to image documents, and the use of free text captioning to enhance retrieval performance (Yang *et al.*, 2001a; 2001b).

Original historical paper works of calligraphy comprise valuable legacy of civilization to mankind. They are fragile, and cannot be turned over and over again by many different people. A well-known protection method is restricting the access only to a few researchers. In order to widely share them with the general public in the CADAL project, many famous paper works are digitalized and published on the Portal to CADAL. The key challenge is how to

* Project supported by the National Natural Science Foundation of China (Nos. 60272031 and 90412014) and the China-America Digital Academic Library Project

manage large digitized calligraphy images to offer fast browsing and retrieval services. There are ways of character-to-image conversion, such as those in (Sclaroff and Pentland, 1995), but there is no existing technique to convert calligraphy character images to computer font. Optical Character Recognition (OCR) does not work for such character images because of their deformation. Since most people are interested in the art of the beautiful styles of calligraphy character rather than the meaning of the character, a simple way is to treat them just as they are images without recognizing them like OCR does. In this paper, we will present the techniques related to Chinese calligraphy character retrieval.

Increased attention is being paid to providing additional visual methods of presenting the search results, especially when users need assistance in refining their search strategies. Visualization of results can aid in browsing as well as interpretation. In (http://www.inxight.com) three specialized types of visualizations (relationship, trend and temporal visualizations) are used in information retrieval applications. In the CADAL digital library, some visualization methods are used to present the search results to specialized group of users, including sign language presentation for deaf users and 3-D visualization of the writing process of the retrieved Chinese calligraphy character for users who learn Chinese calligraphy or appreciate it. The latter will be discussed in this paper.

The remainder of the paper is organized as follows. In Section 2 some techniques for multimedia analysis and video summarization are discussed. Section 3 presents the techniques of low-level features extraction, semantic annotation, image retrieval and some other related techniques. Section 4 discusses the techniques related to Calligraphy character retrieval. The method of 3-D visualization of retrieved Calligraphy character is presented in Section 5. Section 6 concludes the paper with some insight into future work.

MULTIMEDIA ANALYSIS AND SUMMARIZATION

The multimedia analysis and summarization system (as shown in Fig.1) employs technologies for speech recognition and transmission, video (face, etc.) detection and recognition, audio analysis, semantic meaning extraction, video structuring and summarization, etc. Multimedia raw data are mostly non-structured or semi-structured. So, multimedia annotation and structuring should be done in order to provide the services of metadata-based retrieval and structurally based browsing.

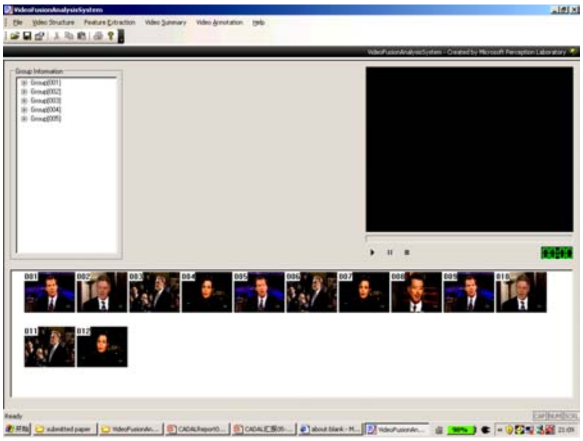


Fig.1 Graphic interface of video analysis system

Video annotation

Given the lexicon, a short video clip can be simply annotated by human interaction. When the video is long, content annotation can benefit from segmenting the video into smaller units. The video segmentation component in our system is based on frame differencing of the color coherence vector histogram, and heuristic rules are designed to make the component robust to flashes and noises. Given shot boundaries, the annotations are assigned to each video shot according to its time order, considering that the semantic information on videos are time-dependent. All the annotation results and descriptions of ontology are stored as MPEG-7 XML files, so that users could use these documents to retrieve and collect video clips conveniently. Fig.2 illustrates the graphic interface of video annotation and metadata generation. Video annotation and metadata generation support video retrieval and semantic analysis.

Video structure generation

As the video stream is composed of thousands of frames, indexing the low-level features of each image frame is ineffective in both space and time and is unnecessary. Moreover, it is unnecessary for users to watch the whole video while browsing and retrieving.

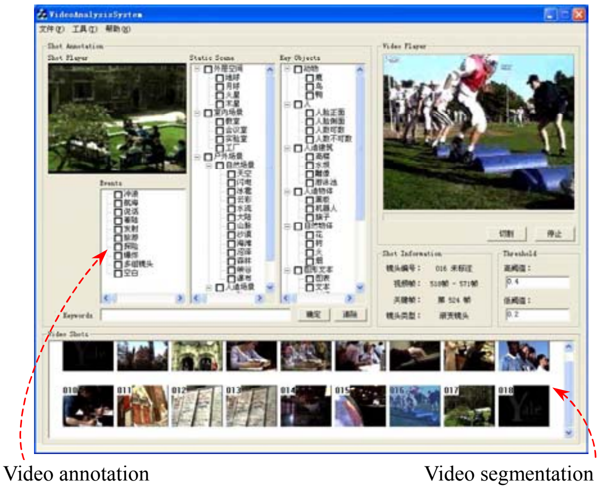


Fig.2 Video annotation and metadata generation

So the method of video structuring is required, by which the video is partitioned into hierarchical structure, the video information of different level is indexed for users' convenient browsing and retrieving.

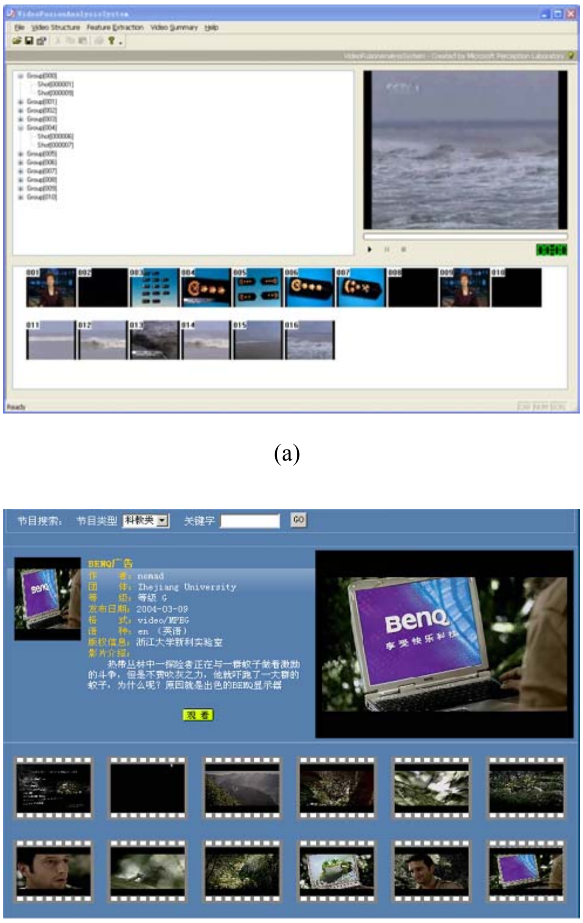
Hierarchical video partition is implemented by automatic shot detection algorithms, i.e. shot cut detection algorithm and shot transition detection algorithm. For shot cut detection, the feature of color coherence vector histogram is extracted; self adaptive threshold and time damping technique is used to ensure the good performance of shot detection. The method of using the first-order derivative of video frame grey means and second-order derivative of grey means is proposed to implement shot transition detection. Fig.3a and Fig.3b show respectively the graphic interface of video structure generation and the interface of structure browsing.

Video summary generation

Automatic summarization can facilitate document selection. Users have become accustomed, in the Web environment, to seeing brief summaries or snapshots of a document's contents, generated on the fly using a variety of summarization techniques. These techniques can be used to add value to retrieval from the digital library.

Video summary is the summary information that mostly covers the semantic meaning of the original video, and is useful for the users' quick browsing and retrieval of the large scale video information. The video summary is generated by compressing the content of the original long video based on the analysis of the video content. As the original video is represented in much more simple mode, video summary generation can greatly save the cost of network bandwidth while users access the video information through network.

The steps of video summary generation are as follows. Firstly, the cut shot and transition shot are detected; then support vector classifier (SVC) is introduced and clustering is done in the high dimensional feature space so that similar shots are clustered into one cluster; lastly the association rule of the clustering set is mined by the video sequence association mining method, and the supporting vector obtained by association rule is regarded as summary. Fig.4 illustrates an example of video summarization.

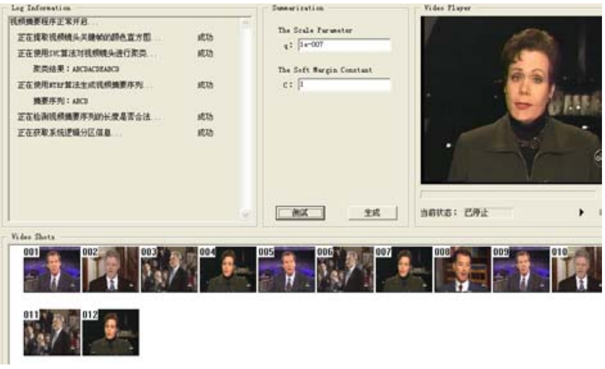


(a)

BENQ广告
作　　者：nomad
团　　体：Zhejiang University
等　　级：等级 G
发布日期：2004-03-09
格　　式：video/MPEG
语　　种：en （英语）
版权信息：浙江大学新科实验室
影片介绍：
热带丛林中一探险者正在与一群蚊子做着激励的斗争，但是不费吹灰之力，他就吓跑了一大群的蚊子，为什么呢？原因就是出色的BENQ显示器

(b)

Fig.3 Video structure generation (a) and video structure browsing (b)



视频摘要程序正常开启...
正在提取视频镜头关键帧的颜色直方图... 成功
正在使用SVC算法对视频镜头进行聚类... 成功
聚类结果：ABCDACDEABCD
正在使用STEP算法生成视频摘要序列... 成功
摘要序列：ABCD
正在检测视频摘要序列的长度是否合法... 成功
正在获取系统逻辑分区信息... 成功

The Scale Parameter
q: 1e-007

The Soft Margin Constant
C: 1

当前状态：已停止

Fig.4 Video summarization generation

IMAGE RETRIEVAL

As the digital library contains millions of unstructured multimedia resources such as images, videos, audios, etc. besides textual information, supporting effective and efficient retrieval of multimedia resources is a challenging problem in the CADAL project. Here we examine the issues related to image retrieval. The retrieval is executed based on the seamless integration of semantic keywords and media-specific low-level features

Low-level features extraction

Semantic features (keyword, annotations) and visual features (color, texture, shape and appearance of object) are selected as the image features. Visual features can be classified into general features and domain-specific features. The former is used to describe the common features of all images, unrelated to the content of the image, such as color, texture and shape; the latter is then constructed based on some hypothesis of the content of the image, related to the corresponding application. Here we select color, texture and shape, which are effective features for image retrieval, as features. The features are as follows: (1) color histogram which are extracted in HSV space, (2) color moment, (3) color coherence vector, (4) self regression texture, (5) Tamura texture coarseness, and (6) Tamura texture orientation.

Image annotation

Multimedia resources are required to be retrieved through high-level semantics, which are traditionally obtained from manual labelling. Well-annotated image collections include Corel image galleries, most museum image collections, the Web archive, etc. However, this approach is liable to be subjective, and requires a huge amount of human effort. As new multimedia resources increase dramatically everyday, an automatic annotation method becomes necessary and important. Classification is a good way to organize large collections of image into categories with different semantic meanings. A semantic skeleton is defined to describe the semantics of an image category, as

SemanticSkeleton=
<ID, Title, KeywordSet, SemanticBlobSet>,

where KeywordSet is the union of all annotated keywords; SemanticBlobSet is a vocabulary of blobs abstractly representing meaningful image regions of the category. SVMs (Support Vector Machines) are used to classify images automatically. Color histogram (in HSV color space) and Tamura texture are used as image features. Statistical learning method is then employed to select the most appropriate keywords for an incoming image on the basis of the annotated image connections. Based on this, we define multimedia metadata as

<ID, Title, Type, Format, KeywordSet, Size, Feature, URI>.

Image retrieval

For feature-based retrieval, the user is required to submit a media object as the query example, and the results are retrieved based on the similarity of low-level features.

Due to the gap between the semantic feature and low-level feature, good retrieval performance based on visual features cannot be ensured. To avoid the problem, relevance feedback is introduced to bridge the gap between the semantic features and low-level features.

Relevance feedback is the interactive process between user and retrieval system, which is the process that the initial query is updated according to the evaluation of the current retrieval results so as to optimize the retrieval results. Due to the gap between semantic and low-level features, the users' evaluation of the retrieval results is used as basis for helping future retrieval. How to reasonably and effectively express the evaluation is the primary role that relevance feedback plays.

We implemented a multimedia search engine, Octopus, which provides peer index and relevance feedback to avoid the gap between the semantics and low-level features, according to the intuitive and simple idea that the semantic concept is hidden in each image and the semantic concept appears apparently in the relation between the image and the other images. Here peer index (each image is indexed by the other related images) is proposed and applied in the image retrieval. In addition, the learning strategy of automatic construction of peer index from users' relevance feedback is proposed in the system too.

Fig.5 shows the interface of the image retrieval search engine in the Portal to CADAL. Fig.6 shows the retrieval results, given a keyword query "flower". Fig.7 shows the retrieval results, given an image query and the further results through relevance feedback.

CHINESE CALLIGRAPHY CHARACTER RETRIEVAL

In terms of calligraphy character, key issues for retrieval are feature extraction and similarity computation. Feature extraction is to obtain discriminative features such as shape to represent calligraphy character image. We retrieve relevant images according to the similarity matching cost. There are many research works on handling shape features effectively (Sclaroff and Pentland, 1995; Mokhtarian *et al.*, 1996; Celenk and Shao, 1999). A multi-scale skeleton-based invariant feature representation is proposed as a shape representation (Ogniewicz, 1994; Telea *et al.*, 2004). In contrast to their approaches, in our system character complexity and shape, the two kinds of features of the calligraphy character are proposed for the real time retrieval of large calligraphy character image database.

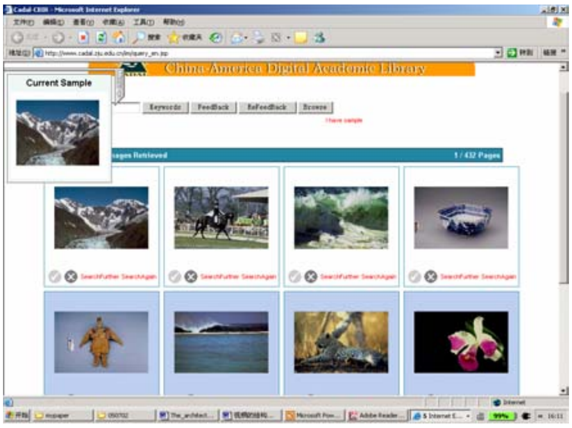


Fig.5 Interface of the image retrieval search engine

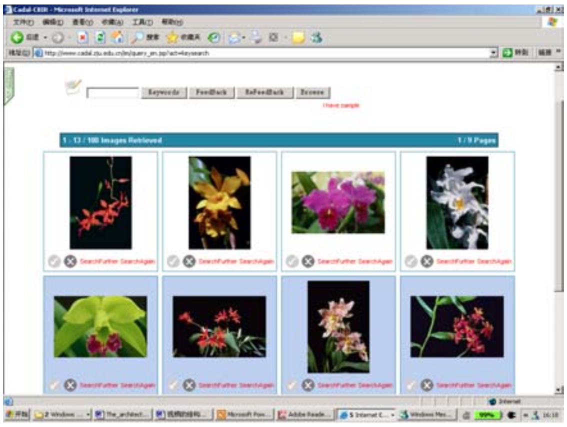
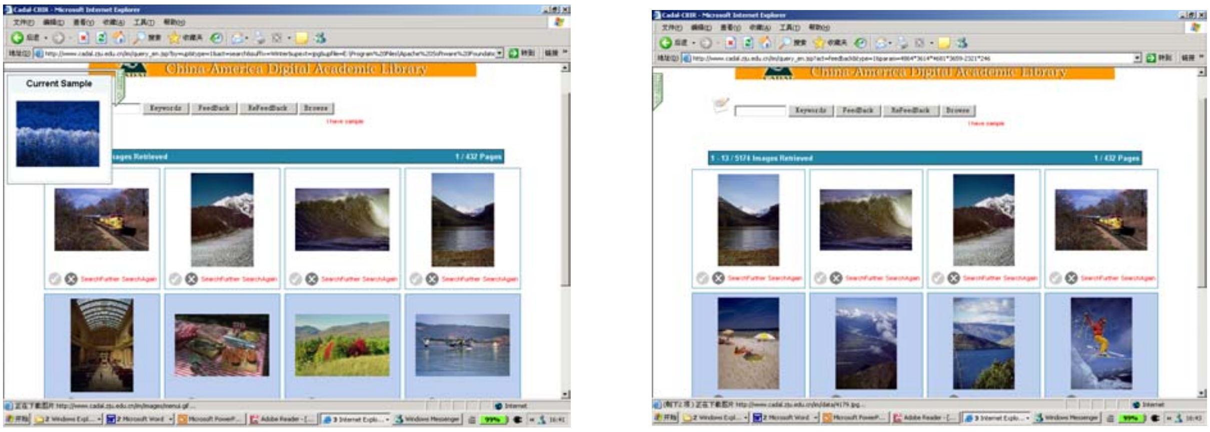


Fig.6 Retrieval results given a keyword query "flower"



(a) (b)

Fig.7 Retrieval results given an image query (a) and further results through relevance feedback (b)

Chinese calligraphy page segmentation

The original calligraphy books, mostly ancient, were scanned at 600 dpi (dots per inch) and kept in DjVu format by the project. The scanned images were smoothed and converted into binary image because the colorful background of the image is not useful in the similarity matching process. Then the scanned original page images are segmented into individual calligraphy characters using minimum-bounding box as introduced in (Manmatha *et al*., 1996). First, the page images are binarized with characters in black and the background in white. Then the pages are cut into columns according to the vertically projecting histogram, with columns continuing to be cut into individual characters. Fig.8 gives an example, showing how a page is cut into individual calligraphy characters. All the characters are normalized in order to keep scale invariant contour information, which is used to represent the calligraphy character in (Celenk and Shao, 1999).

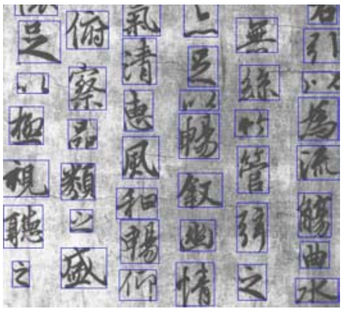


Fig.8 Segmentation example with mark of minimum-bounding box

Features extraction

Here the shape and character complexity of the calligraphy character are extracted.

1. Shape representation

Calligraphic character's shape is represented by their contour points. Shape features are described using approximate points context. Polar coordinates are used to describe the directional relationship of points instead of Cartesian coordinates. For direction, we use 8 equal in degree bins to divide the whole space into 8 directions. And for radius, we use 4 bins using $\log_2 r$. For each point p_i of a given point set composed of n sampling points, we describe its approximate shape context by its relationship with the remaining points in k weighted bins $w_i(k)$.

$$w_i(k)=\#\{q_j\neq p_i:q_j\in bin(k)\} \tag{1}$$

where p_i means ith point which is computed for its point context, k means kth bin of the point. Fig.9 shows an example.

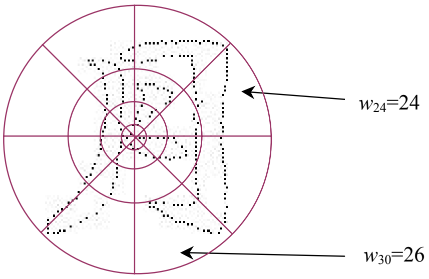


Fig.9 Log-polar bins for point context computing when the point is p_i

2. Calligraphy character complexity

Many calligraphy characters can have same character complexity. Therefore it is not discriminative enough and must combine with other features. We use it as a filter at the beginning to discard the calligraphy character that has no possibility to be similar to the query. Let L be the number of sampled contour points from the query and L_i be the number of sampled contour points from ith candidate image. Then the filter can be written as:

$$1/\alpha\leq L/L_i\leq\alpha \tag{2}$$

where α is the threshold obtained by experience. After filtering, the number of possible candidate is reduced.

Character image retrieval

For the query sample, it can be an existing calligraphic character image imported from the disk. Or, if the user has no query sample initially, it can be sketched, or even typed in using keyboard. The shape feature of this query image is computed for the later matching process. Certain calligraphic character may have more deformation than others. We use feedbacks from user to navigate the user to change query step by step to get the calligraphic character that the user really wants. The retrieval process is as follows:

(1) Compute the values of the character complexity of each calligraphy character in the database.

(2) Normalize the scale size of the query and sample its contour points.

(3) Filter the candidate images by Eq.(2).

(4) Extract the shape feature and employ the shape matching method introduced in (Zhuang *et al.*, 2004) to compute the matching cost for every remaining candidate image and the query.

(5) Rank the results according to the matching cost, and return.

Fig.10 and Fig.11 show respectively the retrieval results for the query "shu" and the graphic interface for browsing the original works and the related information.

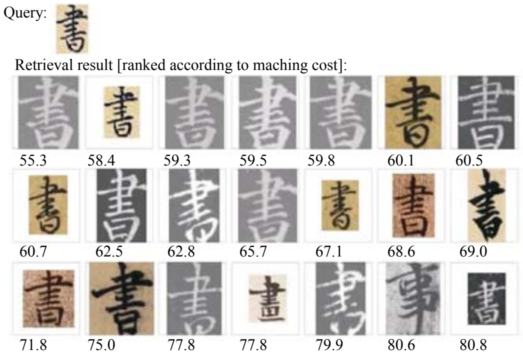


Fig.10 The result of retrieval

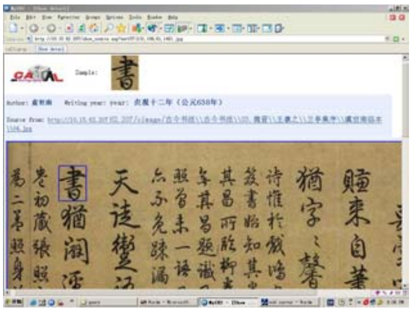


Fig.11 Interface of browsing the original works and the related information

3-D VISUALIZATION OF THE SEARCH RESULTS

Some search results can be interpreted through visualization. Here we take our Chinese calligraphy character retrieval system as a visualization example. In order to help people enjoy the art of calligraphy writing and find how it was written step-by-step, in the system, the writing process of the searched character is animated by 3-D visualization method. Firstly, extract strokes order of an offline Chinese calligraphic character. Then estimate the varied stroke's thickness. Finally, animate the writing process. Users can view the animation of calligraphic writing process by their browsers.

Extraction of skeleton

Thinning algorithm introduced in (Chen *et al.*, 1996) is employed to extract the skeleton of a calligraphic character image. In order to extract the strokes order, the skeletal pixels are categorized into three types (vertex pixel, line pixel and branch pixel) according to its degree.

Extraction of stroke thickness

As every point in the obtained skeleton is discrete, instead of using traditional ellipse (http://en.wikipedia.org/wiki/Stroke_order) to describe stroke thickness, we use the following formula to define the approximate thickness w of each pixel in the original image:

$$w=d+T_b/8d \qquad (3)$$

where d represents the horizontal or vertical distance from the destination skeleton pixel to the ring of which all the points are black, T_b is the number of total black points of the gray ring (as shown in Fig.12).

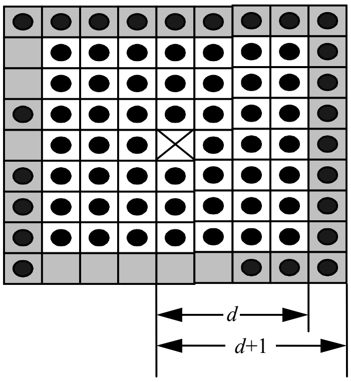


Fig.12 The square representing the stroke of a given pixel

Extraction of strokes order

Spurious vertex and spurious cluster points are removed and the double traced path is marked before strokes order is extracted.

When teaching people how to write a Chinese calligraphy character in the right way, writing rules is very important (Ma *et al.*, 2002; Helena *et al.*, 2000).

Chinese characters have many structure types, such as left-right structure, top-bottom structure, surrounding structure, semi-surrounding structure. These writing rules are useful for extracting strokes order.

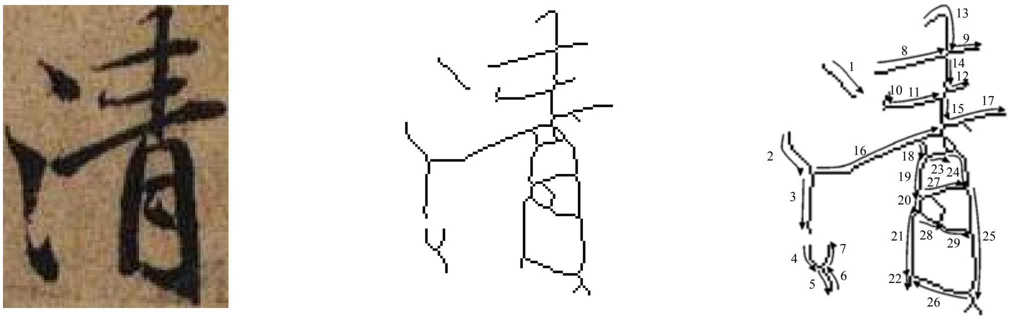
We define the skeleton graph $G=\{g_1, g_2, \ldots, g_n\}$, where g_i ($i\in[1,n]$) represents a sub-graph consisting of strokes connected. The character structure is analyzed by calculating the orientation relationship between each pair of the elements of graph G. The sequence of sub-graph is traversed. During the traversing of the sub-graph, when reaching a cluster point which has more than two paths, the path that has angle close to 180° has higher priority and is selected, because most Chinese calligraphic writings are composed of relatively straight strokes. When intersection point is encountered, the traversal proceeds along the smoothest path, assuming that the calligrapher choose to spend minimum energy when writing a calligraphic character. The idea is especially useful for extracting cursive hand calligraphic writings.

Fig.13a shows an original image of “qing”, written by Wang Xi-zhi, a famous Chinese calligraphist who lived in Jin Dynasty about 1600 years ago. The skeleton after applying the thinning algorithm is shown in Fig.13b. Fig.13c is the result of stroke order extraction. The width distribution of points along the strokes path is shown in Fig.14. Some screen shots of the writing process are shown in Fig.15.

CONCLUSION AND FUTURE WORK

In this paper, we introduced some related novel multimedia retrieval and visualization techniques, such as multimedia analysis and summarization, image retrieval, Chinese calligraphy character retrieval and 3-D visualization of the writing process of the Chinese character. All these techniques related systems have been integrated into the Portal to CADAL (http://www.cadal.zju.edu.cn).

Future work with the Portal to CADAL will proceed in several directions. We will improve the performance of the current services. We will extend our Portal to CADAL to be more complete, and to enable generation of personalized digital libraries. We will integrate the systems related to some other techniques, such as bilingual service, knowledge service and management, multi-modal text to speech conversion and virtual reality, into the Portal to CADAL.



(a) (b) (c)

Fig.13 Original image of “qing” (a); Skeleton of “qing” (b); Extracted strokes order (c)

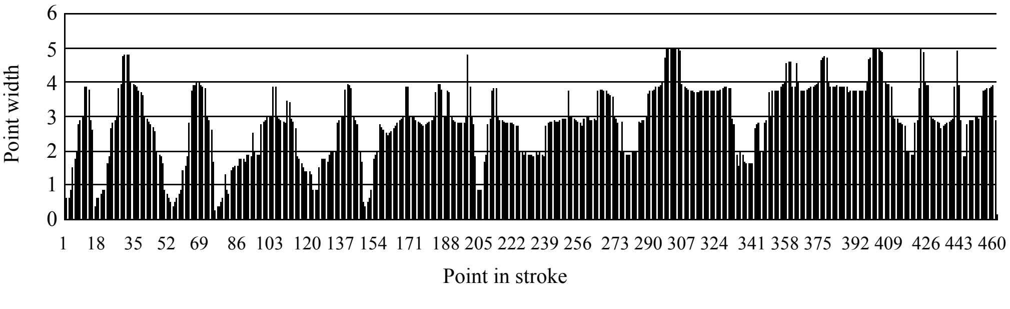


Fig.14 Width distribution of points along the strokes path

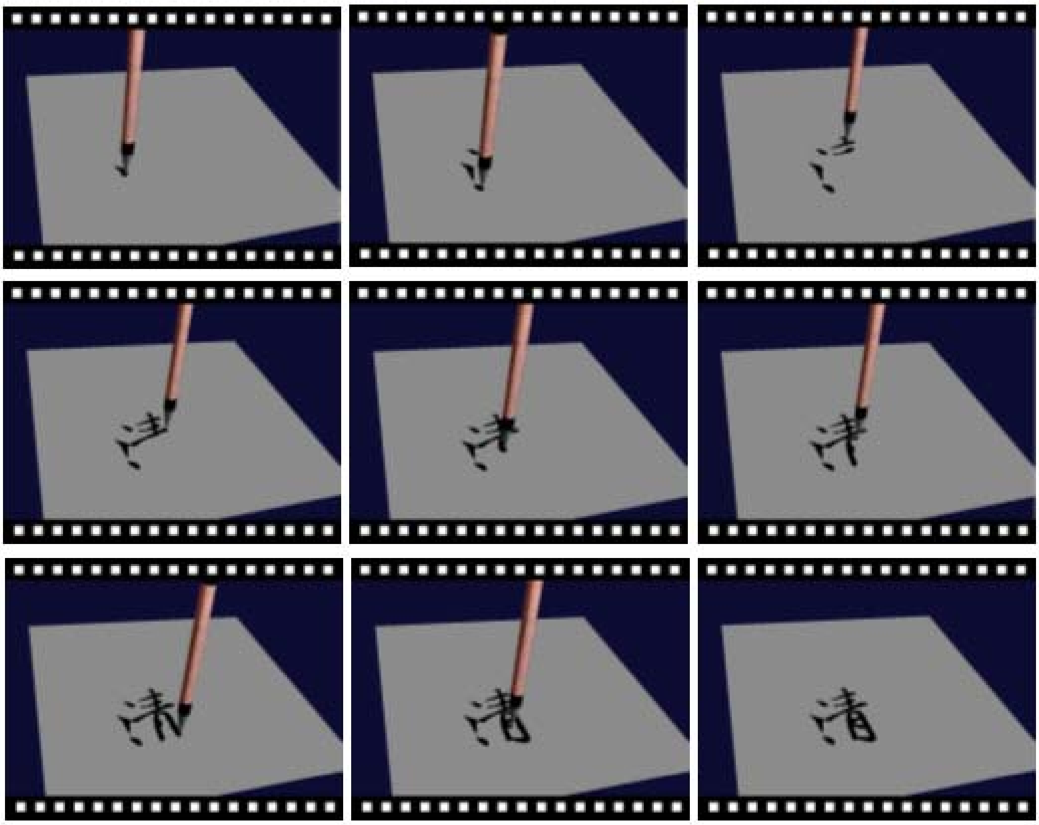


Fig.15 3-D visualization of the writing process of Chinese calligraphic character "qing"

References

Celenk, M., Shao, Y., 1999. Rotation, translation, and scaling invariant color image indexing. *Storage and Retrieval for Image and Video Databases VII. SPIE*, **3656**:623-630.

Chen, Z., Lee, C.W., Cheng, R.H., 1996. Handwritten Chinese Character Analysis and Preclassification Using Stroke Structural Sequence. Proceedings of ICPR'96.

Helena, T., Wong, F., Horace, H.S., 2000. Ip*: Virtual brush: a model-based synthesis of Chinese calligraphy. *Computers & Graphics*, **24**:99-113.

Ma, T.Y., Zhu, H.L., He, B., 2002. Visual C++ Digital Image Processing (Version 2.0). People Post Press, Beijing (in Chinese).

Manmatha, R., Han, C.F., Riseman, E.M., Croft, W.B., 1996. Indexing Handwriting Using Word Matching. Proceedings of the 1st ACM International Conference on Digital Libraries, p.151-159.

Mokhtarian, F., Abbasi, S., Kittler, J., 1996. Robust and Efficient Shape Indexing through Curvature Scale Space. Proc. British Machine Vision Conf., p.545-561.

Ogniewicz, R., 1994. Skeleton-space: A Multiscale Shape Description Combining Region and Boundary Information. CVPR, p.746-751.

Sclaroff, S., Pentland, A., 1995. Model matching for correspondence and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **17**:545-561.

Telea, A., Sminchisescu, C., Dickinson, S., 2004. Optimal Inference for Hierarchical Skeleton Abstraction. IEEE ICPR.

Yang, J., Zhuang, Y.T., Li, Q., 2001a. Search for Multi-modality Data in Digital Libraries. Proceedings of the Second IEEE Pacific Rim Conference on Multimedia (PCM), Beijing, China, p.482-489.

Yang, J., Zhuang, Y.T., Li, Q., 2001b. Multi-Modal Retrieval for Multimedia Digital Libraries: Issues, Architecture, and Mechanisms. Proc. of International Workshop on Multimedia Information Systems (MIS), Capri, Italy, p.81-88.

Zhuang, Y.T., Zhang, X.F., Wu, J.Q., Lu, X.Q., 2004. Retrieval of Chinese Calligraphic Character Image. IEEE 2004 Pacific-Rim Conference on Multimedia, LNCS 3331, p.17-24.