CADAL Digital Library

Wu Jiang-Qin,Zhuang Yue-Ting, Pan Yun-He

College of Computer Science, Zhejiang University

Hangzhou, 310027, P.R.China

***ABSTRACT* — CADAL（China America Digital Academic Library ） is a collaborative project between universities and institutes in China and USA, which aims to provide universal access to large scale digital resources and explore the ways of applying multimedia technologies to digital library. Four distinct characteristics of the resources in CADAL are : (1)the amount of the digital resources including digital books and multimedia for research and education can reach nearly 100 terabyte; (2)various types of media resources are contained, including image, video,3-D model and other types of media resources.(3)both Chinese books and English books are digitalized, so the language is bilingual.(4)Chinese traditional culture resources such as Chinese Calligraphy, are contained in the CADAL resources. Aiming at the above characteristics, in the Portal to CADAL, active services of unified paralleling search for the different types of digital resources, the services of quickly retrieving and structurally browsing of multimedia documents including image, video, the services of bilingual translation and the services related to Chinese traditional culture resources are provided.**

***INDEX TERMS* — CADAL, Chinese Calligraphy, Cross-Media Retrieval, Multimedia Retrieval, Video Summarization**

I. INTRODUCTION

The China-Us Million Book Digital Library(CADAL) is an international cooperation program between China and the US. One million digital books are not the ultimate goal of the project, but it is the first step towards to universal access to human knowledge. There are four challenges for the access of resources of CADAL. Firstly, the amount of the digital resources including digital books and multimedia for research and education can reach 100 terabyte. Secondly, image, video,3-D model and other types of media resources, various types of media resources are included in the CADAL resources. Thirdly, there are two kinds of language digital books. Chinese and English, in the CADAL resources. Fourth and lastly, traditional Chinese culture resources are important part of the CADAL resources. Aiming at the above four challedges, in the Portal to CADAL, unified paralleling searching for the different types of digital resources, quickly retrieving and structurally browsing of multimedia documents including image, video, the services of bilingual translation and the services related to the traditional Chinese culture are provided.

TB volume of various types of digital resources, such as dissertation, ancient minguo book, modern book, minguo journal, English book, drawing, video and illustration are available in the CADAL, which is one of the distinct characteristic of CADAL. So CADAL presents a challenge for the technique of searching resources based on metadata.

There are different types of multimedia resources in the CADAL, so CADAL presents a challenge for the application of multimedia analysis and retrieval techniques. 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 relevance feedback to add information to image documents, and the use of free text captioning to enhance retrieval performance [1,2].

The explosive growth of digital video data in CADAL is calling for more effective approaches to enabling users for quick browsing. How to satisfy users' requirements is a very challenging problem we have to face. To meet the users' needs, we implemented a powerful video structuring and summarization generation system and provided the services of quickly video browsing in the CADAL portal.

A large percentage of data in digital library presents a cross-media feature. Media data of various modalities coexists and the different media data expresses a common semantic topic. So the relationship among the multi-modal objects should be taken into consideration in information retrieval. In order to make the most use of the abundant resources in the digital library, the greatest challenge is to provide a precise and effective retrieval method, so that the user could obtain the expected multi-modal data. So, we proposed a cross-media retrieval algorithm by page ranks

CADAL project integrates Chinese and English books into a digital library. To facilitate inter-communication of people from different cultures, the digital library platform must support bilingual access. In the CADAL portal, if you can input the query word in either English or Chinese, you can get the searched results in both languages. In addition, when you read a book, the system can translate the contents of the page.

Original historical paper works of calligraphy are valuable civilization legacy of mankind and they are important part of CADAL resources. But they are fragile, can't be turned over and over again by many different people. 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 manage large digitized calligraphy images to offer fast browsing and retrieval services. In the CADAL portal, the services of the Chinese calligraphy image character retrieval is provided.

The remainder of the paper is organized as follows. In section 2 some techniques for unified paralleling search for various types of media resources are discussed. Section 3 presents the techniques of low-level features extraction, semantic annotation, image retrieval, video summarization, cross-media retrieval and some other related techniques. The services of bilingual translation are presented in section 4. Section 5discusses in detail the techniques related to Calligraphy character retrieval. Section 6 concludes the paper with some insights into future work.

II. UNIFIED PARALLELING SEARCH

Metadata is a kind of structure data of information resource or data, which is the structure description of information resource. As metadata can describe its corresponding resource by title, authors and so on, a user can find the resource meeting his requirements by searching metadata.

The metadata is extracted for all the resources in the CADAL, so the resources can be searched by metadata for easily locating the required resource.

There are many types of digital resources in the CADAL, including Dissertation, Ancient, Minguo Book, Modern Book, Minguo Journal, English book, drawing, video,dunhuang and illustration. Dublin core metadata is used to describe the million digital books in the CADAL project. Besides DC, metadata corresponding to the other types of multimedia resources are used to describe them. Independent data map is designed for each kind of resource metadata, such as dunhuang, calligraphy, video and journal.

In order to meet the requirements of different users and improve the user's interactive experience, the service of unified parallel searching for the different types of digital resources is provided for users' convenient searching. When users input a query word and choose the resources database they want to access, the portal will search the specified metadata databases corresponding to each specified resources based on the query keyword and the searched results corresponding to each kind of resource will be respectively presented in the form of columns in the unified interface(as shown in Fig.1 and Fig, 2).

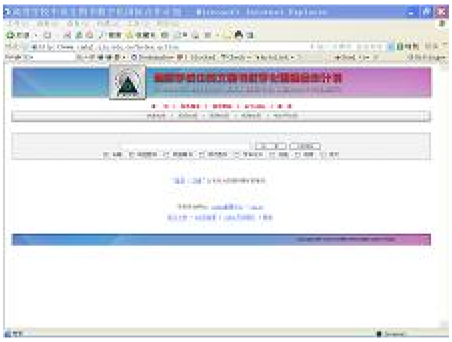

Fig. 1 unified searching interface

Fig. 2 the interface of searched results

III. MULTIMEDIA ANALYSISI AND RETRIEVAL

As the digital library contains unstructured multimedia resources such as images, videos, audios etc besides textual information, effective and efficient analysis and retrieval of image resources is a challenging problem in the CADAL portal. Here we examine the analysis and retrieval issues related to two primary kinds of multimedia, image and video.

A. Image annotation

Image resources are required to be retrieved through high-level semantics, which is traditionally obtained from manual labeling. Well-annotated image collections include Corel image galleries, most museum image collections, the web archive, etc. However, this approach is liable to subjective, and requires a huge amount of human effort. As new image 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. 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.

B．Image retrieval by peer indexing

Here, we presented a new scheme for image indexing — peer indexing — which describe images through semantically relevant peer images. In particular, each image is associated with a two-level peer index, which includes a global peer index describing the "data characteristics" of this image as well as a set of personal peer indexes, each describing the "user characteristics" of an individual user with respect to this specific image. Both types of peer index are learned interactively and incrementally from user feedback information.

Based on two-level image peer indexes, the initial query vectors and similarity metric can be optimized towards a specific query conducted by a specific user before any user intervention by applying a pseudo feedback strategy, a parameter adaptation technique analogous to relevance feedback.

C．Relevance Feedback

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, the retrieval performance based on visual features can not 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 and aid for future retrieval. How to reasonably and effectively express the evaluation is the primary role that relevance feedback will play.

D．Image search engine

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 related other 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.3 illustrates the interface of the image retrieval search engine in the Portal to CADAL, the retrieval results, given an image query and the further results through relevance feedback.

Fig.3 (a) the retrieval results given an image query

Fig.3 (b)the further results through relevance feedback

The retrieval is executed based on the seamless integration of semantic keywords and media-specific low-level features

E．Video analysis System

The explosive growth of digital video data is calling for more effective approaches to indexing and retrieving of video clips based on its content, and enabling users for quick browsing. How to satisfy users' requirements is a very challenging problem we have to face. we implemented a powerful multimedia analysis and summarization system (as shown in Fig.4), employing 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 browsing.

Video Annotation Video Segmentation

Fig.4. Graphic Interface of Video Analysis System

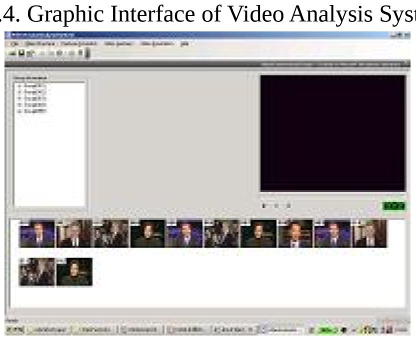

Fig.5 Video Annotation and Metadata Generation

F. 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 of 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. 5 illustrates the graphic interface of video annotation and metadata generation. Video annotation and metadata generation support video retrieval and semantic analysis.

G．Video structure generation

As the video stream is composed of thousands of frames, indexing the low-level features of each image frame is ineffective both in space and time and is unnecessary. Moreover, it is unnecessary for users to watch the whole video while browsing and retrieving. 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 performance of shot detection. The method of using first-order derivative of video frame grey means and second-order derivative of grey means is proposed to implement shot transition detection. Fig. 6(a) and Fig.6 (b) illustrate, respectively, the graphic interface of video structure generation and the interface of structure browsing.

H．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, which 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.7 illustrates an example of video summarization.

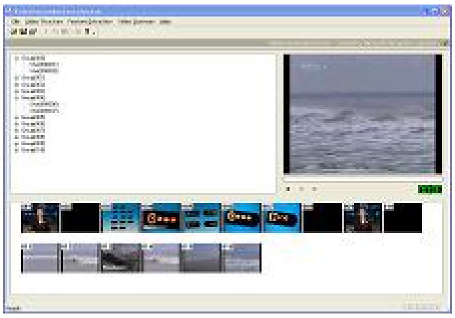

Fig.6 (a) Video Structure Generation

Fig.6 (b) Video Structure Browsing

Fig. 7 Video Summarization Generation

I. *Cross-media retrieval by PageRank*

PageRank is a method for rating web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them [3]. In a word, it is a technique to rank the web pages.

In this section, we introduced the idea of *PageRank* to accomplish cross-media retrieval. the cross-media retrieval algorithm is mainly composed of two parts: query-dependent part and query-independent part.

Graph construction First we need to construct a cross-reference graph *G*.

We define the multimedia document as: *Docu-ment=<DocumentID, URI, ElementSet, LinkSet>*. *DocumentID* is the identification of a multimedia document. *URI* is the uniform resource identification of a document. *ElementSet* is a set of all the media objects in the document. And *LinkSet* is the set of links contained in the document, including the links via which other pages point to the document or the document points to the others.

The steps of constructing the graph: 1) initializes the graph. 2) establish a semantic relationship between the media objects. 3) takes into account the similarity of low-level features. Here, we just find K most content-based similar objects for each object, and regard them as those "similar" with it.

Basis vectors computation Assume that *G* has *N* vertexes. For each of the unit vector x_i *(i=1…N)*, if we regard it as the *preference vector* **u** in equation **V=(1-c) Av+*c* u**

we could compute the corresponding personalized PageRank vector r_i *(i= 1…N)*. *basis vector*s.

PPV computation The two steps above could be regarded as query-independent part. Then the query-dependent part is as follows:

A query could contain one or several examples, and their modalities could be the same or different. Suppose there are *n* query examples of *m* different modalities, with n_i (*i=1...m*) examples of each modality. We use q_{ij} to denote each query example, where *i* and *j* represent the *j*th query example of the *i*th modality. Assume in each modality M_i *(i=1...m)*, there is a similarity function $f_i(x, y)$ *(i=1…m)*, then for each query example q_{ij}, we are able to find the most similar *k* objects (we denote the set of these *k* objects as P_{ij}) of the same modality, each denoted as p_l, where *l* represents the *l*th media object in the **whole** multi-modal dataset (not the *l*th in P_{ij}). Let s_{ijl} denote the similarity of p_l and q_{ij}, then

$$s_{ijl} = \begin{cases} f_i(q_{ij}, p_l), p_l \in P_{ij} \\ 0, p_l \notin P_{ij} \end{cases} \tag{1}$$

if p_l is one of the k most similar objects of q_{ij} then we compute their similarity use the corresponding similarity function, otherwise, if p_l is not similar enough with q_{ij} or they have different modalities, we just set their similarity to 0.

If we give each modality a weight w_i satisfying $\sum_{i=1}^{m} w_i = 1$, then the preference vector **u is**

$$u = \sum_{l=1}^{N} \alpha_l x_l \qquad (2)$$

where

$$\alpha_l = w_i \sum_{j=1}^{n_i} s_{ijl} \Big/ \sum_{l=1}^{N} \sum_{j=1}^{n_i} s_{ijl}$$

on condition that the modality of the *l*th media object is M_i

According to the Linearity Theorem, we could get the *PPV*:

$$v = \sum_{l=1}^{N} \alpha_l r_l \qquad (3)$$

Here v reflects the similarity between the *i*th object and the query. We could select the ones with largest value as the result.

Relevance feedback Relevance feedback is an important way to refine the semantic relationship between the media objects and improve the precision of retrieval. The relevance feedback in our algorithm has two functions: improving the results in one query (short-term) and adjusting the cross-reference graph (long-term).

Improving the results**:** Suppose we get the set of relevant objects R, which has *n* relevant objects of *m* different modalities. The number of objects of each modality is n_i *(i=1…m)*. This time we do not use a similarity function to decide how important each of them is in a *preference vector* **u**, because all of them are considered to be relevant by the user. We simply compute **u** using equation (2) with

$$\alpha_l = w_i / n_i \; (i = 1\ldots N) \qquad (4)$$

Also, we assume the *l*th media object is of modality M_i . In fact, we even do not need to compute **u**. As long as we obtain $\{a_l \mid l = 1..n\}$, we are able to get **v** according to equation (3).

2) Adjusting the graph**:** We add one to the weight between two relevant media objects, and subtract one from the weight between a relevant object and an irrelevant object. If the weight decreases to zero, we delete the link between them. Since the user's feedback is used to update the cross-reference graph, the effect of this feedback is long-term. As the feedback times increases, the relationship between the media objects would become more and more accurate.

After the graph has been adjusted, we need to up-date the *basis vectors* r_i*(i=1…N)*. But we do not need to do this every time the user feeds back, for it is too time-consuming. Instead, it is more reasonable to re-compute the *basis vector*s after several queries have been finished.

IV. BILINGUAL SERVICES

Most current popular digital libraies such as Library of Congress, Chinese National Digital Library and ACM Library, do not provide multi-lingual access interfaces and automatic translation services. As there are both English and Chinese books in CADAL, bilingual services are required for users to access resources in any language. In the CADAL portal, some research work by north technical center on how to carry out the multi-layered bilingual machine translation in English and Chinese books, such as the metadata translation between English and Chinese, the accurate translation of proper nouns such as names for unique individuals, events,or places, the selective translation in a full-text context, the translation of Old Chinese text, and the distributed translation service technique. Technologies and

Fig.8 the interface of the bilingual search

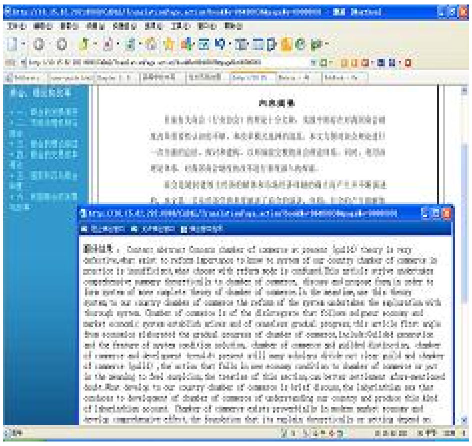

Fig. 9 the translation of contents of a page

prototypes have been developed to accurately translate proper nouns, and selectively translate the metadata of e-books in a full-text context. An online translation service is integrated into the CADAL portal. Users can be directly conducted semantic-based multi-linguistics retrieval of required information in our CADAL digital library.

As shown in Fig 8. the user choose the type of language he needs (English or Chinese), the system will translate the query word and then search and then display the searched results in both English and Chinese language(as shown in Fig. 8). In addition, the services of translation of the contents of the page will be provided when the user read a book(as shown in Fig. 9).

V. THE CHINESE CALLIGRAPHY CHARACTER RETRIEVAL

There are ways of character-to-image conversion, such as [4], but there is no existing technique to convert calligraphy character images to computer font. Optical Character Recognition (OCR) doesn't 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 the CADAL portal, the service of Chinese calligraphy character retrieval is provided.

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. Recently, there are many research works on handling shape features effectively [4][5][6]. A multi-scale skeleton-based invariant feature representation is proposed as a shape representation[7][8]. In contrast to their approaches, in our system character complexity, stroke density and shape, the three kinds of features of the calligraphy character are proposed in order to build index for real time retrieval of large calligraphy character image database.

A. 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 to 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 [9]: First, the page image are binarized with characters in black and the background in white. Then cut the page into columns according to the vertical projecting histogram, and columns continued to be cut into individual characters. Fig. 8 gives an example, showing how a page was 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 as introduced in [6].

B. Features Extraction

Here shape, character complexity and stroke density, the three kinds of features of the calligraphy character are extracted.

- shape representation

Calligraphic character's shape is represented by their contour points. Shape features are described using approximate points context. The polar coordinates is used to describe directional relationship of points instead of the Cartesian coordinates. For direction, we use 8 bins in equal degree size 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 $w_i(k)$.

$$w_i(k) = \#\{q_j \neq p_i : q_j \in bin(k)\} \qquad (5)$$

Where p_i means ith point which is computed for its point context, k means kth bin of the point.

- 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:

$$\frac{1}{\alpha} \leq \frac{L}{L_i} \leq \alpha \qquad (6)$$

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

C. 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 want

the retrieval process is as follows:

(1)Compute the values of the character complexity of the calligraphy character.

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

(3)Filter the candidate images by formulae (6)

(4)Extract the shape feature and employ the shape matching method introduced in [6] to compute the matching cost for every remaining candidate image and the query.

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

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

Fig. 10 the result of retrieval

Fig. 11 interface of browsing the original works

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced some distinguishable services of the CADAL digital library[11], such as the active services of unified paralleling search for the different types of digital resources, the services of quickly retrieving and structurally browsing of multimedia documents including image, video, the services of bilingual translation and the services related to Chinese traditional culture resources. All the services have been accessed by the users from over 70 countries 170.000 times per day.

With the increase of the number of the users and the amount of the resources. 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 be more stable

ACKNOWLEDGEMENT

This research is supported by the National Natural Science Foundation of China (No. 60272031,No. 90412014) and the China-US Million Book Digital Library Project (see http://www.cadal.zju.edu.cn).

REFERENCES

[1] Jun Yang , Yueting Zhuang , Qing Li, Search for Multi-modality Data in Digital Libraries, Proceedings of the Second IEEE Pacific Rim Conference on Multimedia (PCM), pp. 482-489, Beijing, China, 2001.

[2] Jun Yang, Yueting Zhuang, Qing Li, "Multi-Modal Retrieval for Multimedia Digital Libraries: Issues, Architecture, and Mechanisms", Proc. of International Workshop on Multimedia Information Systems (MIS), pp. 81-88, Capri, Italy, 2001.

[3] Y.Wu, Q. Tian, T. S. Huang. Discriminant-EM Algorithm with Application to Image Retrieval. *IEEE Con-ference of Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, 2000, pp. 1222-1227

[4] Sclaroff, S., Pentland, A.: Model matching for correspondence and recognition. IEEE Trans.Pattern Analysis and Machine Intelligence 17 (1995) 545–561

[5] Mokhtarian, F., Abbasi, S., Kittler, J.: Robust and efficient shape indexing through curvature scale space. Proc. British Machine Vision Conf. (1996) 545–561

[6] Celenk, M., Shao, Y.: Rotation, translation, and scaling invariant color image indexing. Storage and Retrieval for Image and Video Databases VII. SPIE 3656 (1999) 623–630

[7] Ogniewicz, R.: Skeleton-space: a multiscale shape description combining region and boundary information. CVPR (1994) 746–751

[8] Telea, A., Sminchisescu, C., Dickinson, S.: Optimal Inference for Hierarchical Skeleton Abstraction. IEEE ICPR (2004)

[9] R. Manmatha, Chengfeng Han,E. M. Riseman,W. B. Croft, Indexing handwriting using word matching, Proceedings of the 1st ACM international conference on Digital libraries, pp.151 – 159,1996.

[10] Wei Wang, Jiong Yang, Richard Muntz, PK-tree: a spatial index structure for high dimensional point data. In: Tanaka K, Ghandeharizadeh S, Kambayashi Y eds.Information Organization and Database. Boston/Dordrecht/London: Kluwer Academic Publishers, pp: 281~293,2000.

[11] http://www.cadal.zju.edu.cn