A Web-Based Cooperation and Retrieval Model of Character Images for Ancient Chinese Character Research

Xuedong Tian[✉], Songqiang Yang, Xuesha Jia, Fang Yang,
and Chong Zhang

College of Mathematics and Computer Science,
Hebei University, Baoding 071002, China
txdinfo@sina.com

Abstract. The character images are the kernel in the research work of ancient Chinese characters. Nowadays, more and more optical sensing device such as scanners and digital cameras are used for transforming ancient Chinese books into digital images for character research instead of paper documents to overcome the difficulties resulted from the rarity and fragility of ancient documents. How to cooperate the images with researchers in web environment becomes a key topic for the higher efficiency and quality of research work. In this paper, considering the requirements in research work of ancient Chinese characters, a web-based cooperation and retrieval model of character images is constructed which consists of several modules including ancient Chinese image management and retrieval, research work cooperation, research conclusion data management and so on. The key techniques employed in this system are discussed. Firstly, the cooperation mechanism of ancient Chinese character images was designed to make character images cooperate with researchers to avoid the occurrence of the situations that a character image is analyzed by several researchers simultaneously or not assigned to any personal for long time. Secondly, the method of integrated image-text arrangement was constructed to adapt to the special situation of ancient Chinese character research that their study conclusion records often contain the mixture of text and images. Thirdly, the conclusion record sorting algorithm by Chinese character radicals was proposed which could arrange the records according to various kinds of dictionary radical orders to meet the needs of research works. Finally, the global and local retrieval method of ancient Chinese character images is proposed for researchers to search the similar character images in database with the global or local area features of character images. The experimental results show that the system constructed with the proposed model has effective assisting function for the research work of ancient Chinese characters.

Keywords: Cooperation and retrieval model · Ancient Chinese character research · Character image · Image retrieval · Image-text integrated arrangement · Conclusion record sorting

1 Introduction

The research on ancient Chinese characters is a meaningful work for the digitalization of ancient documents and the promotion of popularization and dissemination of Chinese civilization.

In a large scale research project in this field, researchers are often confronted with large amounts of characters to be studied. To identify the attributes of an ancient Chinese character including its pattern, pronunciation and meaning, researchers must apply their knowledge about ancient Chinese characters sufficiently, read a lot of related references and discuss with other researchers frequently. Meanwhile, it is also necessary for scientific management among researchers, the research task and the research resources including studied character images and related references. No doubt, employing information technology could bring many conveniences and improve the efficiency for these activities.

The number of Chinese characters is hard to be determined. At present, many ancient Chinese characters have still not been included in existing coded character set, which brings many problems when they are processed in computers. As we all know, the characters in computers are stored in code mode for transmitting and processing conveniently. When a character is to be displayed or printed, the corresponding character pattern is loaded by operating system according to its code. So, a character beyond the coded character set could only be displayed or printed in image mode, which will result in the larger memory consuming and compatibility issues with normal text. Moreover, these character images could not be searched with normal text retrieval techniques, they could only be found out with the help of more complex technology of image retrieval. Unfortunately, most of ancient Chinese characters to be researched are not coded and have to be treated as images. These will bring many problems when they need processed together with ordinary text which consists of coded characters.

Through the aforementioned analysis, it is necessary to research and develop a targeted assisting model for ancient Chinese character research.

The related theory and technology of constructing a managing model for assisting the research work of ancient Chinese characters based on network are related to the theory and technology in many fields including CSCW (Computer Supported Cooperative Work), [1] network, database management system, image processing and retrieval. In detail, the model building relates to website construction, ancient Chinese character resource management and cooperation, image-text integrated arrangement, ancient Chinese character sorting according to different dictionary radicals, ancient Chinese character image retrieval, and so on.

The theory of CSCW has been researched for many years which provide a basic support for our model construction. Its object is to construct a computer based system with which people could cooperate with each other to accomplish a common task cooperatively [1]. Rama and Bishop compared the CSCW groupware system including three commercial systems and four academic systems and designed a set of multidimensional criteria for comparing CSCW systems [2]. Penichet et al. proposed a classification method of CSCW system based on logical principles in a flexible and appropriate way [3]. Chen discussed the key problems of cooperative platform system

including system hierarchy structure, user interface, consistency maintenance, concurrent control, access control and record management. A prototype system RITIS (Real-time Image and Text Interactive System) composed of clients and servers with centralized and peer-to-peer structure was constructed. It supports the real-time image and text interaction in Internet environment and multi-user interface of WYSIWIS [4].

Image-text integrated arrangement is to organize text and images in a layout concertedly. To realize this object, the spatial information of images and characters must be recorded and utilized [5, 6]. Compared with the technology of text information processing, the image-text integrated arrangement is more complex in sides of input, edition, display and output [7–12]. Yang and Cheng designed a scheme to realize image-text mixed arrangement based on XML. They employed the design pattern of MVC to separate text and its view for minimizing the coupling degree among modules and improving the extensibility of system with java language. Because of the platform independent attribute of java and XML, the proposed prototype system has better flexibility [5]. Lu proposed a method to realize the storage technology of image-text mixed arranging documents and their online edition using the open source server control FreeTextBox and ASPJpeg of ASP.net [6]. Zhang et al. put forward a B/S mode based question bank management system. Through using DSOFramer container and the edition function of Word, they realized the input, edition and composition of test paper containing image-text integrated questions [7]. Fan discussed the method of image-text mixed arrangement of test paper in which not only text but also images frequently appear. He classified the layouts into three types called non-image layouts, single image layouts and multiple images layouts. VB and SQL server are used to process and manage the three types of layouts respectively [8]. Zhang and Chen studied the collection, storage and extraction methods of test paper which contain text, formulas, tables and images. They use Delphi as the developing tool and realize the import and export of test papers [9].

In the field of image clustering and retrieval, two types of strategies called text-based method (TBIR) and content-based method (CBIR) are adopted for obtaining required images from image library [13–20]. Yang researches on image clustering and its application in image retrieval. Through analyzing the existing image clustering features and algorithms, an image retrieval system based on image clustering is designed. The images in library are clustered with AP algorithm and an image index is built firstly. Then, the sample image is searched in the index to find the corresponding class. And the following image matching operation is fulfilled only in this class [13]. Xie studied on the topic of image clustering and retrieval. To solve the problem existing in traditional image clustering algorithms, an image clustering method based on MRF is proposed which transforms the clustering task into energy minimization process. And a local image retrieval method is designed with graph cutting mode [14]. Zhuang et al. proposed a novel method of retrieving Chinese calligraphic characters. The images of Chinese calligraphic characters are matched by the feature of approximate point correspondence algorithm. After the contour points are extracted, the approximate point correspondence is computed and the matching operation of character images is run according to their accumulated matching cost [15]. Zhuang et al. put forward a retrieval method of Chinese calligraphic manuscript images based on probabilistic indexing structure called PMF-Tree (Probabilistic Multiple-Feature-Tree).

Integrated features are used in retrieval such as contour points of character images, character styles and types. The characteristic of this method is that users are allowed to select one of above features as retrieval components [16]. Chen proposed an image retrieval method based on integrated features of global statistic feature and local bitmap feature. The mean-variance of RGB values of images are calculated as the global feature. Then, the image is divided into sub areas to get the mean value with binaryzation processing as the local feature. Finally, image retrieval program is run with the combination of the global and local features [17]. Kong et al. design a semi-supervised image retrieval method. The characteristic points of an image are extracted with improved Harris algorithm. The image is divided into the regions of interest and the color and texture features are extracted. Then, the semantic relation between the image and its class is established through semi-supervised learning in image feature space. Finally, the similarity between images and class centers are computed [18].

The theory and technology on the construction of network stations assisting for research work have become mature; the details of them will not be discussed here.

The above work laid the foundation for our research and developing work. In this paper, considering the requirements of ancient Chinese character research, a web-based cooperation and retrieval model of character images for ancient Chinese character research is constructed which is composed of several modules including ancient Chinese image management and retrieval, research work cooperation, research conclusion and reference management. The key techniques employed in this system are discussed such as the character image cooperation mechanism, image-text integrated arrangement, conclusion data sorting of ancient Chinese character research according to Chinese characters radicals, global and local retrieval of ancient Chinese character images, and so on.

The paper is organized as follows. Section 2 outlines the overall architecture and functions of the model. In Sect. 3, the key techniques employed in the model are analyzed and introduced. The experimental result is introduced and analyzed in Sect. 4. Finally, conclusions and the further work are discussed.

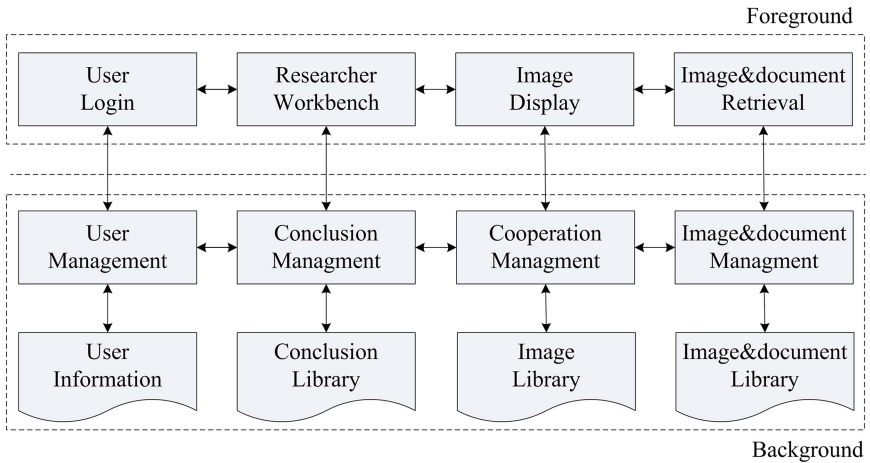## 2   Architecture of the Cooperation and Retrieval Model

The object of the model is to realize the cooperation management among character images to be studied, researchers and the records of research conclusions. Meanwhile, it provides the image and document retrieval service for Chinese character researchers in the process of research work. The architecture of the model is shown in Fig. 1.

The input data of the model is the images of single ancient Chinese character.

Ancient books are digitalized with optical sensing devices (scanners or digital cameras) to form the layout images firstly. Then, page layout analysis and character image segmentation program is employed to segment these layout images into a series of single character images supplied to ancient Chinese character researchers.

The output data of the system is the records of research conclusion data which includes the pattern, pronunciation, meaning and so on of each character image.

Based on the requirement analysis to the ancient Chinese character researchers, the design principles of the model are as follows.

Foreground

User Login	Researcher Workbench	Image Display	Image&document Retrieval

User Management	Conclusion Managment	Cooperation Managment	Image&document Managment

User Information	Conclusion Library	Image Library	Image&document Library

Background

Fig. 1. The architecture of the cooperation and retrieval model

Principle 1. Uniqueness principle. Each image is to be given a unique key code when it is storied into the library of character images and should be allocated to only one researcher for studying.

This could avoid the occurrence of the situation that one character image is assigned to more than one researcher at any time which will result in the confusions of conclusions.

Principle 2. Hierarchy principle. The users of the model are divided into different levels with different authority according to their roles in research work.

Users of the platform with different authorities have different operation scopes, which could effectively avoid the fault operations to the research data.

Principle 3. Independence principle. The research conclusions of a researcher about a character image could only be modified by himself. Other people could give suggestions to him rather than change his research records.

This item is to protect the data of research conclusions from modified by other people rather than original researcher of the character image himself.

Principle 4. Compatibility principle. No matter coded characters or images of no coded characters, the system could organize them normally with the mode of image-text integrated arrangement.

The image-text integrated arrangement problem exists not only in the display operation of research data but also the import and export of the research records in database. So, it is necessary to design a special structure to tackle these problems to ensure the normal use of conclusion records by researchers.

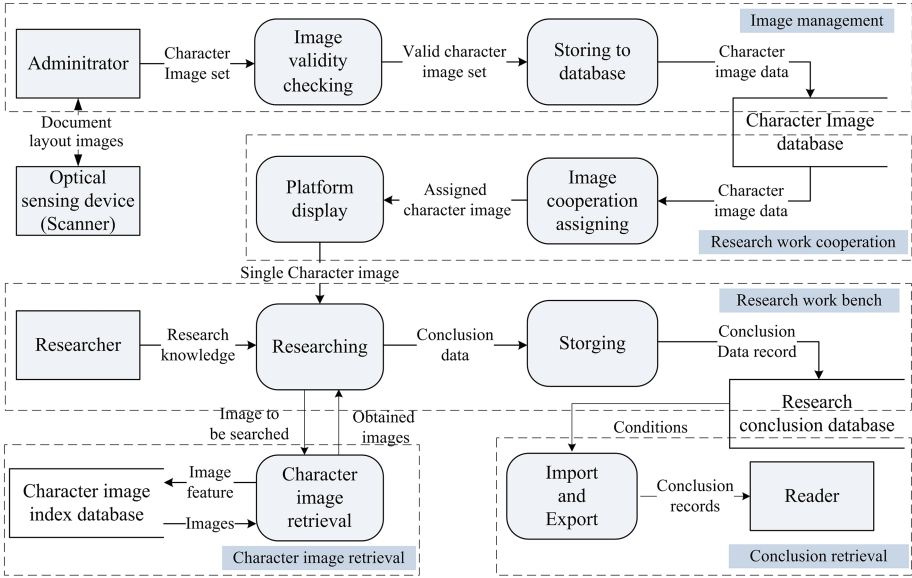The data flow diagram of the model is shown in Fig. 2.

Fig. 2. Data flow diagram of the model

3   Key Techniques in the Model

3.1   Cooperation Mechanism of Research Task

According to the principles of the relationship among researchers, images and research conclusion records, a field controlling operation must be done in data library of different image elements.

Assume ResearchState to be a field in ancient Chinese character image library, SelectID to be a field in researcher library and ExpertNumber to be a field in research conclusion library. The definition of the field value in cooperation is shown in Table 1.

Table 1. Cooperation attributes in library.

Library/Field	Value/Statements
Image/ResearchState	**1**: Current image has not been researched
	2: Current image is being researched
	3: Current image has been researched
Researcher/SelectID	**-1**: The researcher has not allotted an image
	Primary Key of image: The researcher is researching the image
Research conclusion/ExpertNumber	**Primary Key of researcher**: The record could only be modified with the researcher

Through proper setting operations to the semaphores shown in Table 1, the assigning principles of research resource could be abided to ensure the normally running of research work.

3.2   Image-Text Integrated Arrangement

Image-text integrated arrangement mainly consists of three modules: image-text integrated display, edition and their import and export of research data.

3.2.1   Image-Text Integrated Display

In this module, Literal control is employed to realize the image-text integrated display. The coded characters and the tags of image addresses are stored in character strings of Literal. The text attributes of Literal control is linked to Bind ("literal"). The image-text integrated display is shown in Algorithm 1.

Algorithm 1. Image-text integrated display

Input: The content of research conclusions to be displayed in database.

Output: The display data of the content**.**

  (1) Get the content to be displayed from the field of database table.
     (a) Get the coded characters.
     (b) Get the tags of image addresses.
  (2) Store the characters and the tags of image addresses to be displayed into the Literal control.
  (3) Make the text attributes of Literal control linked to Bind ("literal").

3.2.2   Image-Text Integrated Edition

This module includes the functions of character input and the insertion of character images. Iframe control is utilized to realize the image-text integrated edition which creates an inline frame including a document.

When a character image is inserted, it is stored into database and renamed uniformly. The numbers of character images are recorded and corresponding image labels stored with characters publicly are established.

The algorithm of inserting images into text is shown in Algorithm 2.

Algorithm 2. Image-text integrated edition

Input: The data of character codes and images to be edited.

Output: The image-text integrated data.

  (1) Input the content into iframe control.
    (a) Input characters into iframe control.
    (b) Insert character images into iframe control correspondingly.
  (2) Store the character images into server and rename them uniformly.
  (3) Record the number of character images and store the corresponding image labels.
  (4) Store the characters and the tags of image addresses into the iframe control.

3.2.3   Image-Text Integrated Import and Export

The attribute of the library field in which the labels of characters and images of literal is stored is set to "nvarchar". When we export them into word document, an operation of transforming the html labels to images is fulfilled.

When the data in word documents is imported, the detected images are renamed and stored into database. The locations of the images are replaced with the address links and stored into database. The attributes of the corresponding library fields are set to "nvarchar". The image-text integrated export is shown in Algorithm 3.

Algorithm 3. Image-text integrated export

Input: The research conclusion items.

Output: Word documents including the research conclusion.

(1) Read the research conclusion items.
(2) If the format of the item is image, go to step (3); else go to step (4).
(3) Insert the image. The algorithm terminates.
(4) If the format of the item is not Literal, go to step (5); else go to step (6).
(5) Output the text. The algorithm terminates.
(6) Read the characters of the literal one by one.
(7) If not the end of the string, go to step (8), else the algorithm terminates.
(8) If the character is the tag of the image, go to step (9), otherwise go to step (10).
(9) Convert the tag of the image into the image and output the image, go to step (6).
    (10) Output the text, go to step (6).

The image-text integrated import is shown in Algorithm 4.

Algorithm 4. Image-text integrated import

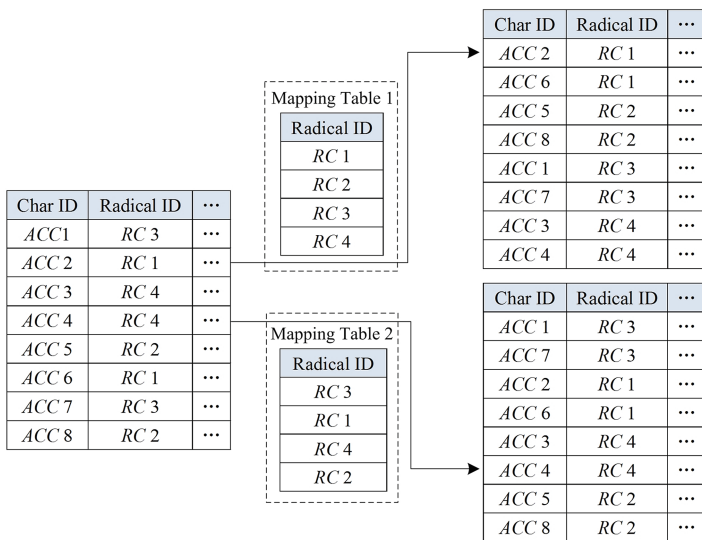
Input: Word documents including the research conclusion.

Output: The research conclusion item.

(1) Read the data in the imported word document.
(2) If not the end of the string, go to step (3), else go to step (7).
(3) If the format of the character is image, go to step (4), else go to step (6).
(4) Rename the images by the server time and store them into server.
(5) The locations of the images are replaced with the address links. Go to step (1).
(6) Store the text of the word into corresponding field of database. Go to step (1).
(7) Integrate the data and store them into database. The algorithm terminates.

3.3   Research Conclusion Records Sorting by Radical Order

There exist various kinds of radical sets with their corresponding orders related to ancient Chinese characters. So it is necessary to establish a mapping table to realize the sorting function of conclusion records according to different radical sets. The mapping operation of radical sorting of library records is shown in Fig. 3.

Char ID	Radical ID	⋯
ACC 2	RC 1	⋯
ACC 6	RC 1	⋯
ACC 5	RC 2	⋯
ACC 8	RC 2	⋯
ACC 1	RC 3	⋯
ACC 7	RC 3	⋯
ACC 3	RC 4	⋯
ACC 4	RC 4	⋯

Mapping Table 1

Radical ID
RC 1
RC 2
RC 3
RC 4

Char ID	Radical ID	⋯
ACC1	RC 3	⋯
ACC 2	RC 1	⋯
ACC 3	RC 4	⋯
ACC 4	RC 4	⋯
ACC 5	RC 2	⋯
ACC 6	RC 1	⋯
ACC 7	RC 3	⋯
ACC 8	RC 2	⋯

Mapping Table 2

Radical ID
RC 3
RC 1
RC 4
RC 2

Char ID	Radical ID	⋯
ACC 1	RC 3	⋯
ACC 7	RC 3	⋯
ACC 2	RC 1	⋯
ACC 6	RC 1	⋯
ACC 3	RC 4	⋯
ACC 4	RC 4	⋯
ACC 5	RC 2	⋯
ACC 8	RC 2	⋯

Fig. 3. Mapping operation of radical sorting of research conclusion records

Assume ACC_i to be the key value in conclusion table, RC_j to be the key value of the radical of ACC_i in radical table, $RS(RC_j)$ to be the key value of RC_j in current mapping table.

Then, the series number of record ACC_i in sorted list could be calculated as

$$RS(ACC_i) = \left(\sum_{k=1}^{n} N_{\text{PR}k} \big|_{RS(RC_k) < RS(RC_j)} \right)$$
$$+ \left(\sum_{l=1}^{m} N_{\text{EQ}l} \bigg|_{\substack{RS(RC_l)=RS(RC_j) \\ ACC_l < ACC_j}} \right) + 1 \tag{1}$$

where $N_{\text{PR}k}\big|_{RS(RC_k) < RS(RC_j)}$ is the number of the records whose $RS(RC_k)$ is less than $RS(RC_j)$, $N_{\text{EQ}l}\bigg|_{\substack{RS(RC_l)=RS(RC_j) \\ ACC_l < ACC_j}}$ is the number of the records whose $RS(RC_l)$ is equal to $RS(RC_j)$ and $ACC_l < ACC_j$.

3.4   Ancient Chinese Character Image Retrieval

A retrieval algorithm of character images is specially designed in the model to assist researchers to find the local or global similar character images in database. It contains not only the traditional image retrieval functions oriented on the whole area or partial area of a character image, but also a new image searching style called image retrieval in symmetrical areas of character images for searching radicals in Chinese characters.

Assume A to be the area of an ancient Chinese character image which is composed of the sub area a_{ij}

$$A = \left(a_{ij}\right)^{T}, (i = 0, 1, \ldots, m - 1; \; j = 0, 1, \ldots, n - 1) \tag{2}$$

where m is the row number and n is the column number of meshes divided according to the principle of elastic mesh [19] within the character image A as shown in Fig. 4(a). The directional line elements feature [20] is extracted in sub areas a_{ij} to form corresponding feature vector as shown in Fig. 4(b).
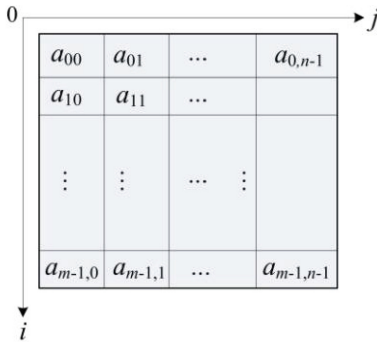
(a) The elastic mesh of character image A
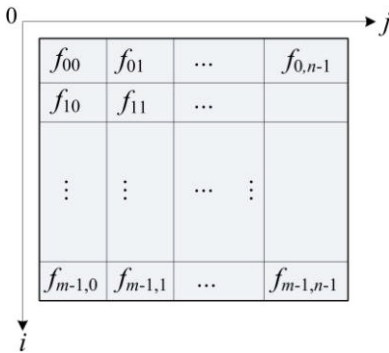
(b) The directional line elements feature F extracted from character image A

Fig. 4. Area division and feature extraction of character image A

$$F = \left(f_{ij}\right)^{T}, (i = 0, 1, \ldots, m - 1; \; j = 0, 1, \ldots, n - 1) \tag{3}$$

where f_{ij} consists of four directional components.

To improve the efficiency of image retrieval, a hierarchy strategy is employed in which a character image A is clustered into sub clusters previously according to the typical areas A_U, A_D, A_L, A_R, A_C and A_W defined by the structural characteristics of ancient Chinese characters.
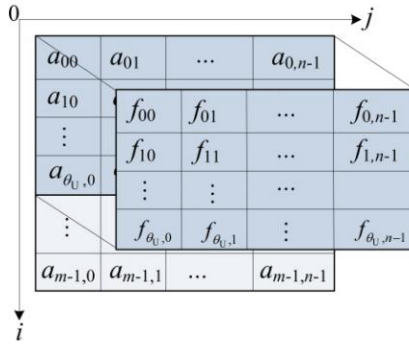
The local areas A_U and A_D in vertical are defined as:

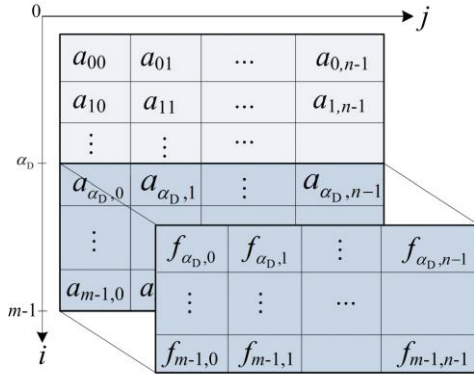$$A_U = \bigcup_{i=0}^{\theta_U} \bigcup_{j=0}^{n-1} a_{ij} \tag{4}$$

$$A_D = \bigcup_{i=\alpha_D}^{m-1} \bigcup_{j=0}^{n-1} a_{ij} \tag{5}$$

where θ_U is the coordinate of the vertical margin of A_U, α_D is the coordinate of the horizontal margin of A_D.

The pixel areas of A_U and A_D in character image is shown in Fig. 5(a) and (b).

(a) Area and feature of A_U in character image A

(b) Area and feature of A_D in character image A

Fig. 5. Area division and feature extraction of A_U and A_D in character image A

The local areas A_L and A_R in horizontal are defined as:

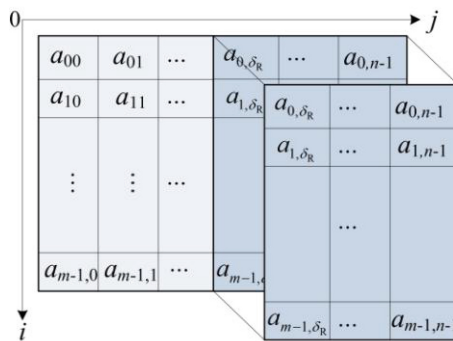$$A_L = \bigcup_{i=0}^{m-1} \bigcup_{j=0}^{\beta_L} a_{ij} \tag{6}$$

$$A_R = \bigcup_{i=0}^{m-1} \bigcup_{j=\delta_R}^{n-1} a_{ij} \tag{7}$$

where β_L is the coordinate of the vertical margin of A_L, δ_R is the coordinate of the horizontal margin of A_R.

The pixel area of A_L and A_R in character image is shown in Fig. 6(a) and (b).
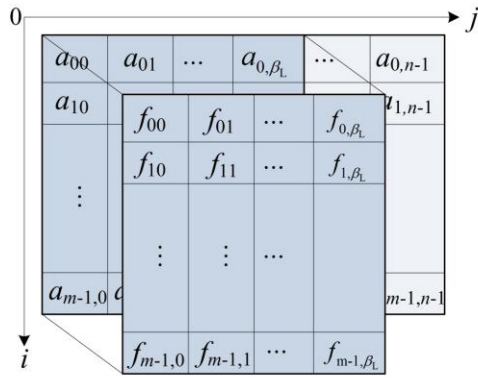
(a) Area and feature of A_L in character image A

(b) Area and feature of A_R in character image A

Fig. 6. Area division and feature extraction of character image A

The local areas A_C and A_W are defined as:

$$A_C = \bigcup_{i=\gamma_{C1}}^{\gamma_{C2}} \bigcup_{j=\mu_{C1}}^{\mu_{C2}} a_{ij} \tag{8}$$

$$A_W = \bigcup_{i=0;}^{m-1} \bigcup_{i=0}^{m-1} a_{ij} \tag{9}$$

where γ_{C1} and γ_{C2} is the coordinate of the vertical margin of A_C, μ_{C1} and μ_{C2} is the coordinate of the horizontal margin of A_C.

The pixel area of A_C in character image is shown in Fig. 7, while A_W includes the whole area of character image.
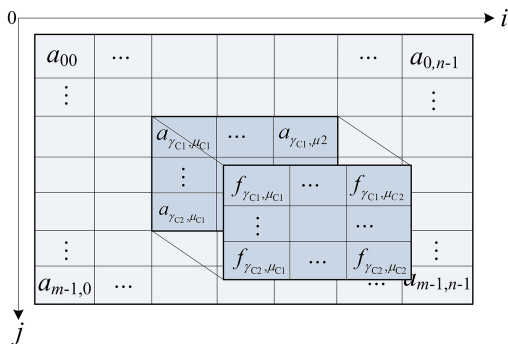

Fig. 7. Area division and feature extraction of A_C in character image A

The clustering operation is fulfilled according to the typical areas. The clustered classes generated by typical areas are C_U, C_D, C_L, C_R, C_C and C_W correspondingly.

When an area A_X is drawn by user within the character image area, it will be matched within the above typical areas A_U, A_D, A_L, A_R, A_C and A_W to find out a best one A_k as the guidance of the following retrieval operation.

$$A_k = \max[(A_X \cap A_i)] \tag{10}$$

where $A_i = A_U, A_D, A_L, A_R, A_C$.

The character image will be searched only in the sub cluster C_k corresponding to A_k based on the feature extracted from A_X. The similarities are measured by the feature distance in $F = (f_{ij})^T$ of A_X and all samples in C_k. The selected character images are returned to the user according to the value of similarities in ascending order.

When a symmetrical retrieval is required, the character image is searched not only in its corresponding classes but also its symmetrical area A_{XS} in character image in vertical direction or horizontal direction such as A_U coupled with A_D, A_L coupled with

A_R. The clustering and retrieval algorithm of ancient Chinese character image is shown in Algorithm 5.

Algorithm 5. Ancient Chinese character image clustering and retrieval

---

Input: Character images.

Output: Matched character images sorted by similarity.

(1) Extracting feature from images and clustering them according to A_U, A_D, A_L, A_R, A_C and A_W correspondingly.
(2) Remark the clustering class ID of each character image.
(3) Obtain A_X.and calculate A_k.
(4) Extracting feature of A_k.
(5) Search the character images in corresponding clustering class C_k.
(6) Sort the retrieval results according to the similarity in descending and display them. The algorithm terminates

---

4   Experimental Result and Analysis

A web-based cooperation and retrieval system of ancient Chinese character images for research with the proposed model in this paper is implemented and utilized in a research project of ancient Chinese characters. The system employs Visual Studio 2008 as the development tool, SQL server 2005 as the storing database, and ASP.NET applied to the system as website frame work.

When a user registers to the system, he will be assigned an authority according to his research task. The setting of authority of the system is shown in Table 2.

Table 2. Setting of user authority.

No.	Level	Authority
1	I	① Research work assigning ② Character discrimination ③ Character research ④ Character image retrieval ⑤ Research conclusion management ⑥ Document retrieval
2	II	① Character discrimination ② Character research ③ Character image retrieval ④ Document retrieval
3	III	① Character image retrieval ② Document retrieval

A researcher studies the character image assigned to him in the system workbench. He could use the image retrieval function of system to find the similar characters he need locally or globally in database.

The research conclusions of characters could be written by researcher in the corresponding text edition boxes and stored into library. If a researcher needs to write an image of a character, he could use image-text integrated edition mode.

The efficiency of image-text integrated operations is decided by the ratio of the number of coded characters and character images. In our system, there are 1521 records, about 112437 characters in the table of conclusion results. The number of coded characters and character images is shown in Table 3.

Table 3. The number of characters and images in image-text integrated arrangement.

| No.	Field ID	Character number	Image number
1	3	654	868
2	4	42080	66
3	6	5739	45
4	7	8403	57

The records of conclusion data of character research stored in database could be displayed in different orders according to the needs of specialists. Table 4 shows several kinds of mapping table for radical sorting.

Table 4. Radical mapping table.

No.	Name	Radical serial number		
		Pinyin	Stroke number	*Kangxi*
1	zhe2	201	6	5
2	bing1	10	22	15
3	shi2	138	8	24
4	da4	29	37	37
5	shan1	132	43	46
6	ya3	176	66	92
7	er3	39	120	128
8	che1	18	150	159
9	qing1	124	168	174
10	ma2	97	200	200

In side of ancient Chinese character image retrieval, we scanned 3176 character images of ancient books for experiment with scanner. Through pre-processing of character images, the feature of clustering and retrieval are extracted. All character images are clustered into six clustering classes according to the definition of A_U, A_D, A_L, A_R, A_C and A_W. Each character image is remark with a clustering class ID.
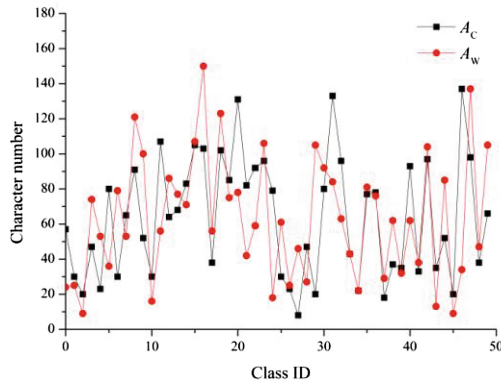
The distribution of character images in different class is shown in Fig. 8.
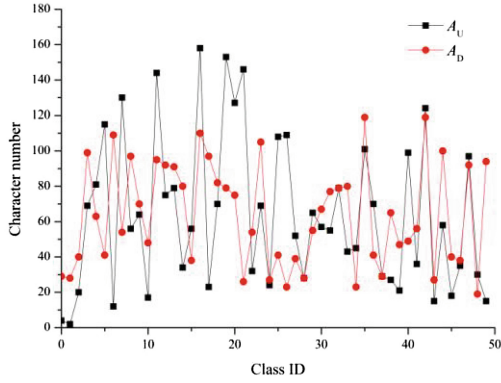
(a) A_U and A_D
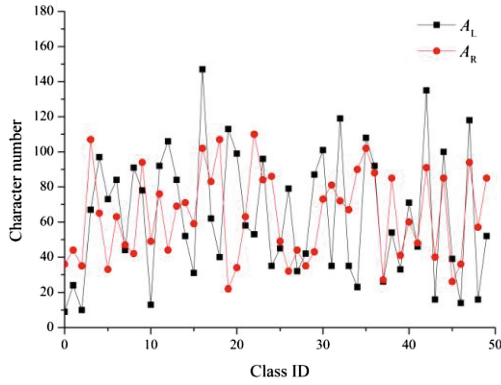
(b) A_L and A_R

(c) A_C and A_W

Fig. 8. The distribution of character images in different clusters. (a) The number of character images of every clusters in local area clustering of A_U and A_D. (b) The number of character images of every clusters in local area clustering of A_L and A_R. (c) The number of character images of every clusters in local area clustering of A_C and A_W.

From Fig. 8 we can see that the numbers of character image samples have similar values among classes. This means the high efficiency when image retrieval is done in different classes.

When a searching area is obtained from screen area drawn by user on the character image, the feature of character image is extracted and the retrieval operation is only fulfilled in the corresponding clustering class of the current character image.

5 Conclusion

In this paper, a web-based cooperation and retrieval model of ancient Chinese character images is proposed for assisting ancient Chinese character research work. It includes the functions of ancient Chinese image management and retrieval, research work cooperation, research conclusion data management, and so on. Because of the characteristics of ancient Chinese characters, many problems need to be solved specially in the constructing process of the system. The key techniques developed in this system, such as image cooperation mechanism of ancient Chinese characters, the method of integrated image-text arrangement, the algorithm of character sorting according to Chinese character radicals and the global and local retrieval method of ancient Chinese character images are discussed. The proposed system realized the designing objects of the model including resource management, and research work cooperation. It is helpful for the improvement of the efficiency of character research work.

Our further work is to improve the performance and complete the function of the system. Firstly, in ancient Chinese character image retrieval, we will optimize the algorithm to obtain higher searching speed. This work needs to select more efficient features of character images and corresponding clustering and retrieval algorithms. Secondly, we will develop the recommendation function of research references. This relies on the knowledge arrangement of researchers in ancient Chinese character research work. Thirdly, better user-friendly menus will be developed. No doubt, these improvements will enhance the assisting ability of our model for a better service to the research work of ancient Chinese character research.

References

1. Shi, M.L., Xiang, Y., Yang, G.X.: Theory and Application of Computer Supported Cooperative Work. Publishing House of Electronics Industry, Beijing (2000)
2. Rama, J., Bishop, J.: 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on it Research in Developing Countries, pp. 198–205. South African Institute for Computer Scientists and Information Technologists (2006)
3. Penichet, V.M.R., Marin, I., Gallud, J.A., Lozano, M.D., Tesoriero, R.: Electronic Notes Theor. Comput. Sci. **168** (2007)

4. Chen, X.T.: M D Paper of Harbin University of Science and Technology, Harbin (2004)
5. Yang, Y.G., Cheng, W.Q.: Comput. Appl. Softw. **24** (2007)
6. Lu, Z.Y.: Comput. Program. Skills Maint. (2013)
7. Zhang, J., Fan, Q., Mao, Z.L., Xu, J.Q.: J. Yunnan Univ. (Nat. Sci.) **32** (2010)
8. Fan, J.H.: China Educ. Info (2006)
9. Zhang, W., Chen, S.Q.: Comput. Mod. (2006)
10. Luo, L.X., Zheng, Z.X.: Sci. Technol. Inf. (2007)
11. Huang, K.: Sci. Mosaic (2009)
12. Sun, Q.L., Liu, X.: Comput. Program. Skills Maint. (2009)
13. Yang, C.H.: M D Paper of Nanjing Normal University, Nanjing (2013)
14. Xie, N.Q.: Image clustering and retrieval based on MRF models. M D Paper of Hefei University of Technology, Hefei (2013)
15. Zhuang, Y., Zhang, X., Wu, J., Lu, X.: Retrieval of Chinese calligraphic character image. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3331, pp. 17–24. Springer, Heidelberg (2004). doi:10.1007/978-3-540-30541-5_3
16. Zhuang, Y., Jiang, N., Hu, H., Hu, H., Jiang, G., Yuan, C.: Probabilistic and interactive retrieval of Chinese calligraphic character images based on multiple features. In: Yu, J.X., Kim, M.H., Unland, R. (eds.) DASFAA 2011. LNCS, vol. 6587, pp. 300–310. Springer, Heidelberg (2011). doi:10.1007/978-3-642-20149-3_23
17. Chen, Y.F., Xue, Q.F., Chen, F.F., Chen, X.D.: J. Changchun Educ. Inst. **29** (2013)
18. Kong, C., Zhang, H.X., Liu, L.: J. Shandong Univ. (Eng. Sci.) **44** (2014)
19. Liu, W., Zhu, N.B., He, H.Z., Li, D.X., Sun, F.J.: J. Chin. Inf. Process. **21** (2007)
20. Wang, H., Ding, X.Q., Halmurat: J. Tsinghua Univ. (Sci. Technol.) **44** (2013)