Document Set Redundancy Compression Method Using Template Differential

Ping Yu, You Yang
School of Computer and Information Science, Chongqing Normal University, Chongqing, China
40465742@qq.com, youyung@yahoo.com

***Abstract*—Document image information systems are used more and more in government. Much redundant information in the document existed in such systems. That implies the research on the compression method based on the page-page statistical features is quite significant. Set Redundancy Compression (SRC) is such a technique that reduces the total entropy of the whole image set by utilizing the image page's similarity. Compression-based Template Differential (CTD) is an improved SRC. The similar image set is constructed by the document template. The coding performance is improved by adding the template image into the Min-Max Differential (MMD) coding/decoding model. It proves theoretically that CTD's coding performance is higher than MMD's. It is demonstrated by experiments that both the CTD and MMD are benefit to increase the compression ratio of image set, however CTD increases more than MMD.**

***Index Terms*—document image; set redundancy; image compression; template differential**

I. Introduction

More and more government resource documents are digital or electronic form. This utilization breaks the time and space restrictions of paper. The digital form document makes the information independent from the paper. Low cost and high security are the advantages of digital documents. But the digitalization of paper documents needs the support of image processing. The quality of processing is critical for the document image information system (DIIS). In such image processing, the image compression is one of the most important process. –compression quality and storage space are the most significant factors of image compression. On one hand, the compression quality relates to the information's first attribute--truthfulness. This attribute determines the utility value of DIIS. On the other hand, the storage space affects not only the cost of hardware, but also the transformation efficiency.

In DIIS application, image compression makes use of not only the page-inner redundancy, but also the page-page redundancy. That is, the statistical features of whole image set should be explored, instead of the individual page's features. The goal of DIIS is to reduce the entropy of whole image set. For example, in industrial and commercial enterprises, a registered document information system exists much redundancy between page and page. Some application sheets and registered sheets have certain format, much information in these sheets are similar. The differences between these sheets are distinguished by the enterprise. About 40% documents could be regarded as similar images in the industrial and commercial DIIS [1]. In the system, there are over 200, 000 enterprises, and there are averagely 90 pages per enterprise. Every enterprise has such pages that company registered, such as corporate representative autobiography table, articles of incorporation. The amount of these similar pages in different enterprises is over 6,400,000.

The Set Redundancy Compression (SRC) technique could be used to deal with such similar images. SRC was proposed firstly by Karadimitriou K. in 1996, Louisiana National University. The concepts of similar image and set redundancy were defined then. The compression techniques included as MMD (Min-Max Differential) [2, 3], MMP (Min-Max Predictive) [2, 4] and Centroid Method [5] were proposed in succession. These three methods used the statistical features of page-page. SRC could decrease the entropy of whole set. More the similarity existed in page-page, more the entropy decreased.

Later, SRC was studied by some researchers. In 2002, HCM (Hybrid Compression Model) was proposed by Jiann-Der Lee [6]. HCM divided the medical image into sub-images with the area growing firstly. Then centroid method was employed for every sub-image. The compression ratio of HCM was 5.6%~134.9% and higher than the traditional centroid method. In 2006, the performance of MMD, MMP and Centroid Method was compared by Ait-Auodia S. etc [7]. Medical images such as CT (Computer Tomography) and MR (Magnetic Resonance) were tested in their experiments. The results demonstrated that the compression ratio employed by SRC was higher Among Bizp2, Gzip, Huffman, RAR and ZIP, SRC was most helpful for Huffman, and least helpful for Bzip. In the same year, MST(Minimum Spanning Tree) that supported effectively CBIR(Content-Based Image Retrieval) was employed to build the similar image set by Nielsen C.and SRC techniques was employed to compress the images. [8].

Although the compression of document was researched for years, it's difficult to improve the compression ratio significantly because of the document complexity content. The MRC (mixed raster content) standard (ITU-T T.44) specified a framework for document compression which can dramatically improve the compression/quality tradeoff [9]. MRC compression is to separate foreground

doi:10.4304/jsw.7.5.1052-1058

from background layers in the document. The key point of this method is the image segmentation. Zaghetto presented a pre/post-processing algorithm for MRC compression [10]. In order to send digital document images in the internet, Immure describes a compression technique for printed document images and string matching method on scanned document images [11]. Although this method could compress 100 pages of document in gray-scale at 300dpi, and the size is around one megabyte, there are two different points between this method and the proposed approach. First， this method can only deal with the inner-page redundancy. Second, the goal of the proposed approach is for information system, not only for Web application.

But, it's difficult to find the references about the application of SRC in DIIS. Because large set redundancy existed in DIIS, it's possible to employ SRC to document similar images. What we can do is to build the similar image set, and construct an appropriate SRC coder/decoder. We realize that the document similar image sets building is much more difficult than the medical similar image sets. In fact, the digitalizing methods and medical methods are different. Medical image is single modality image, and document image is multi-modality image. For the large amount of document, the document images are manufactured. In the paper, the building method of document similar image set is not discussed in detail. Our main goal is to construct the SRC coder/decoder, which is suitable for DIIS.

According to the introduction above, the essential concepts of SRC will be described in section II. In section III, how to define a template from a similar image set will be introduced. In section IV, a novel SRC we called CTD (compression based on template differential) will be proposed. Then, the experiments and analysis of CTD will be followed in section V. Finally, the conclusion will be summarized.

II. Set Redundancy Compression

Image redundancy not only exists in page-inner, but also in page-page. After the introduction of similar image, a page-page compression technique called SRC will be introduced.

A Similarity between images

A definition of similar image was given by Karadimitriou [2]. "Similar images" are images that have the following features

a) Similar pixel intensities in the same areas

b) Comparable histograms

c) Similar edge distributions

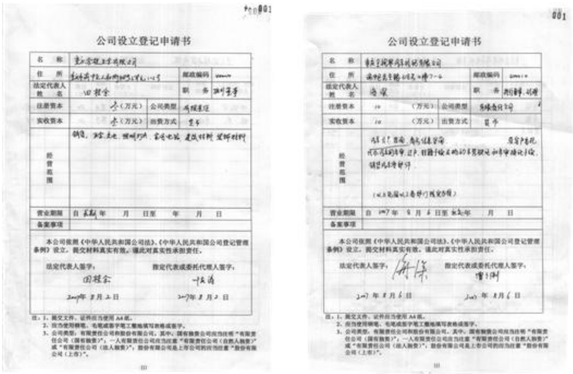
d) Analogous distributions of features

According to the definition above, we could define the similar document images in DIIS: these images in DIIS which have fixed format, or the same page structure. Fig. 1 is an example of the similar images.

By concept above, we could measure the similarity in an image set. Set $f_1 \sim f_N$ the similar images. If these images' similarity is S, $S \in [0,1]$, then the similarity of document image set is: for any position of pixel, there are $S \times 100\%$ images whose pixel gray-level in this position are the same or almost the same. This concept is described as:

$$\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} card\left\{ f \in SSDI \left\| f(x,y) - f_i(x,y) \right| \le t \wedge f_i \in SSDI \right\} \ge S$$

Where, $card$ represents the number of similar document images. t is the threshold of similar image. The concept of the similarity in document image set is illustrated in the following figure.



公司设立登记申请书

公司设立登记申请书

Figure 1. A pair of similar images

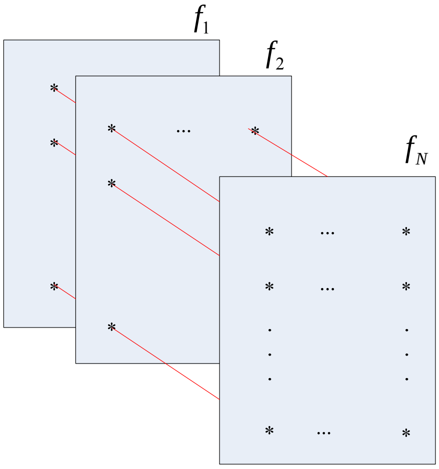


Figure 2. Similar images in an image set

For example, if the similarity of a document image set is S, then we could define the variation of this set V. And

$$V = 1 - S.$$

If $S = 0.6$, averagely, for every pixel in the images, there are 60% images in the set at the same pixel position whose gray-level is equal or almost equal. There are 40% images in the set whose pixels are varied.

B SRC

Such a technique, we called SRC, is the method we reduce the redundancy of similar images. The coder/decoder model was put forward by Karadimitriou K. earlier. The model is illustrated by Figure 3.The model includes two processes, which named set redundancy extraction and individual image compression. The first process is decor relating individual images in the similar image set. The second process is coding the individual image in a certain way. Two meaningful concepts are implicated in the model. One is that =the two processes are independent. The techniques used in the first process would not affect the image compression method in second process. So that we could recognize the first process is a pre-processing procedure of the second. The other concept is that both the processes could decrease the redundancy. The first process decreases the redundancy between the similar images. And the second decreases the redundancy in the individual image. Therefore, the methods based on SRC have another compression way, compared with ordinary individual image compressions.

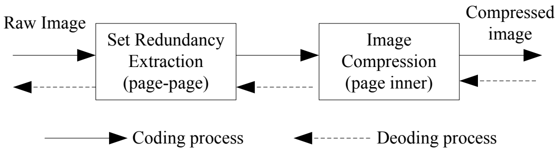

Raw Image
Set Redundancy Extraction (page-page)
Image Compression (page inner)
Compressed image
Coding process
Deoding process

Figure 3. SRC coder/decoder model

The reason that SRC decreases the set redundancy is SRC could reduce the entropy of the image set by utilizing the statistical features between pages to page. Suppose that there is an alphabet table $\{a_0, a_1, ..., a_n\}$, which has $n+1$ symbols. a_k ($1 \le k \le n$) is the symbol whose frequency is S, i.e. $P(a_k) = S$. Then, the sum of the probabilities every symbol has is:

$$1 = \sum_{j=1}^{n} P(a_j) = S + \sum_{j \neq k} P(a_j)$$

According to the Shannon's entropy definition, the entropy H_T of this similar image set could be obtained by following equation.

$$H_T = -S\log S - \sum_{j \neq k} P(a_j) \log P(a_j) \qquad (1)$$

For simple computation, we could suppose that the frequencies of all the alphabets except a_k are equal. That means

$$P(a_j) = V/n, \quad j \neq k.$$

So, (1) could be simplified to (2).

$$H_T = -S\log S - V\log\frac{V}{n} \qquad (2)$$

From (1) and (2), we know that the entropy of the similar image set will decease with the increase of image similarity.

C Max-Min Differential Method

"MIN image", "MAX image" and "Average image" were used in the SRC methods proposed by Karadimitriou K. They represented the statistical features of page-page, and could be calculated by similar image set. For every pixel position, MIN image, MAX image and Average image are minimum, maximum and average among all the similar images in the set. Based on these three images, SRC could reduce the dynamic range of similar image's gray level. So that, the variance of every similar image distributed is decreased. Therefore, the byte counts every pixel needs to execute are less, and the compression is achieved.

MMD is the most important SRC method. The MMD method stores statistical information from a set of similar images in the form of a "min" and a "max" image. To create the "min" image, MMD compares the every pixel position to the pixel value across all images, and chooses the smallest pixel value. Similarly, the "max" image is created by selecting the largest pixel value for every pixel position. Then, MMD processes every image in the set by replacing the original pixel values with the difference from either the "min" or the "max" image (whichever is smaller). This reversible operation reduces the dynamic range of pixel values, so that any standard compression method can be used on the MMD-processed images with improved results.

Figure 4 illustrated the basic idea of this method. Every curve describes an image. The curves for the "min", the "max", and a random image from the set are depicted. The difference values that replace the original image values are shown as dotted lines. Note that the differences are calculated from either the "min" or the "max" curves, depending on which yields the smaller difference value. When the difference is larger than (*max-min*)/2, MMD switches to the other curve. In this way the decoder is synchronized with the encoder, while the smallest possible difference values are selected and used.

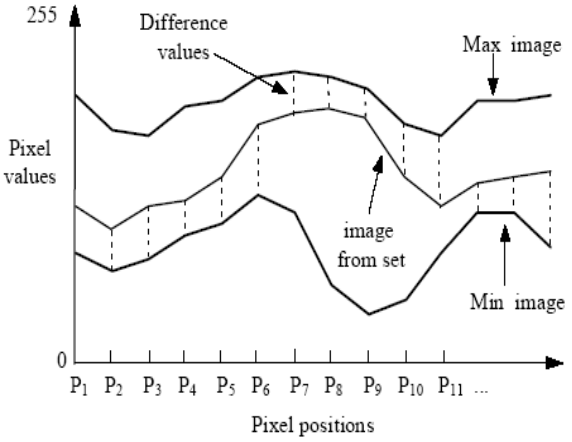


Figure 4. The Min-Max Differential Method

We'll improve this coder/decoder as follows. Let x denote the pixel position, and $value(P_x)$ denote the gray level of pixel poison x in processing image, MIN image and MAX image respectively. The MMD coder and decoder are given by (3) and (4).

$$D_x^{MMD} = \begin{cases} value(P_x) - min_x, & if \ \left(value(P_x) - min_x\right) \\ & \quad < \left(max_x - value(P_x)\right) \\ max_x - value(P_x), & otherwise \end{cases} \qquad (3)$$

$$value(P_x)=\begin{cases} D_x^{MMD}+min_x, & if \ \left(value(P_x)-min_x\right) \\ & <\left(max_x-value(P_x)\right) \\ max_x-D_x^{MMD} & otherwise \end{cases} \quad (4)$$

III. Template Definition

MMD and MMP methods explored the redundancy in page-page by all the similar images in a set. The proposed differential method explored the redundancy by the template. To achieve compression based on content, a template was built to direct the page compression. Based on the template, appropriate method for each sub-image in a page could be utilized to achieve high compression ratio. However, the proposed method is used for an image set. Therefore, the template is the compression direction of all the similar images in the set.

A. Set Redundancy in DIIS

Firstly, the similarity between images in DIIS existed constitutionally. In an industrial and commercial DIIS, registered enterprise is the element information organization. No matter it is logoff or is operating, every enterprise includes the document images such as enterprise license, director material, or official documents, etc. You categorize these documents into 18 types [1]. The most similarity is among the documents between enterprises. For example, the enterprise licenses of the i'th enterprise and the j'th enterprise are similar. Many contents and backgrounds of the license image are the "same". Only the information about enterprise title, license number and address is different. And the similarity is the foundation which we apply on CTD. Additionally, similarity exists also among the documents in the same enterprise. But it's difficult to explore this kind of redundancy.

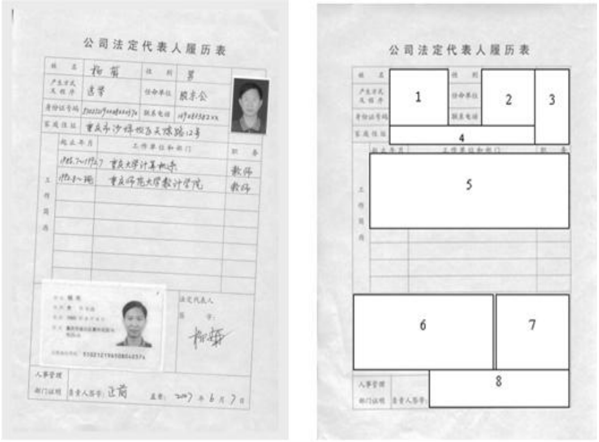
For the similarity in different enterprises, the similar image set could be built by the category template [12]. Template is the reference page or standard page of a category. One template corresponds to an image set. In the set, all the images are similar with each other. The similarity could be the accordant brightness in the same area of the images, or the comparable histogram of the images, or the similar edge distribution, or other feature distribution of the images.

Finally, the storage space could be reduced by SRC application in a DIIS. There is un-conflicting for applying SRC on compressing the similar image set built by template.

B. Template Obtained

Template is important to increase the similarity in a set. If we applied SRC directly on similar document images, then the compressed performance may decreased because of the variation at pixel position. An alternative method is to divide the similar images into two class regions. One class region is the sub-image in which information in the region is non-variable. Another region is the sub-image in which information in the region is variable. For the example of figure 4, these regions in which information comes from artificial writing such as name, photo, ID, seal and signature etc, are class 2; we called ROI (region of interest). Other regions excepted class 2 is class 1, we called SROI (similar ROI). Obviously, the similarity of SROI is higher than ROI. SRC could be applied to SROI. On the other words, SRC is more efficient in SROI than ROI. For the compression of ROI, different compression strategies should be taken according to the attributes of the ROI. Generally, three attributes of ROI could be defined: image, figure and text.

Figure 5(b) is the template of figure 5(a). There are 8 ROI sub-images. They are the main variation parts of similar document image set. Other regions are SROI, which are similar parts of the set. Rectangular was used to divide the image by the two reasons bellow. One reason was that the Chinese document has block structure as rectangular. Another was that the divided regions could independent with no overlap. Because one template corresponds to one similar set, some attributes of template could be defined to describe the similar images features in order to compress the similar images.



(a) Similar image (b) Set template

Figure 5. Relationship between similar image and its set

C. Template Attributes

Template attributes are used mainly to provide priority knowledge for compression. Regions with different attributes should use different methods to compress. Here, a four element vector was used to describe the template attributes. The vector denoted as $T(A_1, A_2, A_3, A_4)$, where T means a template corresponding to similar image set. $A_i\ (i=1,2,3,4)$ is the attribute of the template.

A_1 represents the position information of ROI. The position information is determined by the template but not the similar images. If rectangular is used to divide the template image, then the left-up position coordinates could be used as A_1. For example, $A_1=(502,\ 635)$ in figure 4, it means that the minimum row and the minimum column of ROI 1 in template T are 502, 635 pixels respectively.

A_2 represents the size information of ROI. The size information is determined by the width and height of the rectangular. For example, $A_2=(476,\ 465)$ in figure 4, it means the width and height of ROI 2 were 476, 465 pixels respectively. Based on the combination of A_1 and A_2, the ROI position in template could be calculated. This calculated result was furthermore used in image segment of similar images.

A_3 represents the type information of ROI. The value domain of A_3 is $\{TZ, IZ, GZ\}$, where TZ, IZ and GZ represented text region, image region and figure region respectively. This attribute implied the direction of ROI compression. ROI with different A_3 value should take advantage of different compression method. If the value of A_3 equals to IZ, techniques such as plane fitting with inter-block prediction[13], linked significant tree (LST) wavelet coding method[14], and the adaptive fractal image compression[15] could be employed to finish the compression tasks. If the value of A_3 equals to TZ, techniques such as reference [16] mentioned could be used.

A_4 represents the image registration information of ROI. The value domain of A_4 was {edge, contour, line, interest points, texture}. There are two main usages of A_4. One is describing the features of base image—template image. Another is implying the extraction method of feature set from float image—similar image. Whenever we know A_4, it's much easy to extract the features such as feature type and reference position.

An example of template attributes was illustrated in table I. Four coordinate values were defined. They represented the four table corner positions of the template. The goal of these attributes defined was applied in image registration. It should be noted that different attributes A_i ($i = 1, 2, \cdots, n$) should be defined with different template images. The rule to define A_i was to satisfy the requirements of image compression.

TABLE I. AN EXAMPLE OF THE TEMPLATE ATTRIBUTES

ROI	A_1	A_2	A_3	A_4
ROI1	(502,635)	(500,465)	TZ	(502,305) (3220,305) (499,2203) (3217,2208)
ROI2	(500,1434)	(476,465)	TZ	
ROI3	(499,1910)	(295,598)	IZ	
⋮	⋮	⋮	⋮	
ROI8	(2940,983)	(1225,277)	TZ	

IV. SRC USING TEMPLATE DIFFERENTIAL

The idea compressed based on template differential comes from the gray level reducing solution. The less the pixel level has, the less bit the coder needs. MMD finishes its gray level reducing by comparing the processing image with the MIN image and the MAX image. CTD achieves the gray level reducing not only by comparing the processing image with the MIN image and the MAX image, but also with the template image. In fact, template image is a similar image in the set. Every pixel value of template image is between the value of MIN and MAX.

From the idea above, we defined CTD's coder as:

$$D_x^{CTD} = \min\{\max_x - value(P_x), value(P_x) - \max_x, |value(P_x) - value(T_x)|\} \quad (5)$$

In which, x represented the pixel position. D_x^{CTD} representes the differential value produced by CTD coder. $value(P_x)$ and $value(T_x)$ were the gray level of processing image and template image in position x respectively. The decoder was defined as:

$$value(P_x) = \begin{cases} \max_x - D_x^{CTD}, & c1 \\ \min_x + D_x^{CTD}, & c2 \\ value(T_x \pm D_x^{CTD}), & otherwise \end{cases} \quad (6)$$

Where conditions $c1$ and $c2$ are:

$$\max_x - value(P_x) = D_x^{CTD}$$

and

$$value(P_x) - \min_x = D_x^{CTD}.$$

When $value(T_x) > value(P_x)$, the operator "±" in (6) got minus "—". Otherwise, it got "+".

Figure 6 illustrated the essential concept of CTD. The abscissa denotes the pixel position; the ordinate denotes the gray level of every pixel. It's clear that the template image was between the MIN image and MAX image, or it's closer to the processing image.

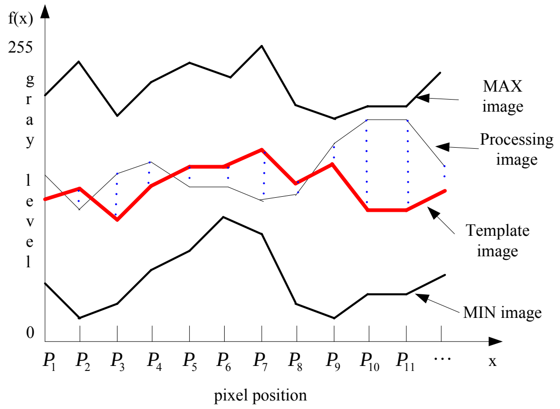


Figure 6. Concept diagram of CTD

From Figure 6, we have the inequality bellow.

$$\min_x \le value(T_x) \le \max_x$$

Comparing CTD's coder in (5) with MMD's coder in (3), we get

$$D_x^{CTD} \le D_x^{MMD} \quad (7)$$

This equation demonstrated that the performance of CTD is higher than MMD.

V. EXPERIMENTS

Documents images from an industrial and commercial DIIS were selected to do experiments. These images come from different enterprises, but they have the same category. A similar image set was built by these images. That is

$$T = \left\{ I_j \middle| I_j \in DIIS \wedge \left[I_j \right] = "XX" \right\}$$

Where, T was a similar image set, I_j represented the similar image in DIIS, $\left[I_j \right]$ represented the attribute of I_j.

In T, I_j may come from the different scale enterprises, or may come from the different area enterprises, or may come from different registered time enterprises.

The results compressed by CTD were shown in table II. In the table, two solutions were described. They were the results of PSNR=33.9115 and PSNR=29.5835. When PSNR=29.5835, the document image size didn't exceed 25KB, which was the condition of practical application in a DIIS [1].

The compression ratio improved (RI) was defined to evaluate the performances of SRC. It was

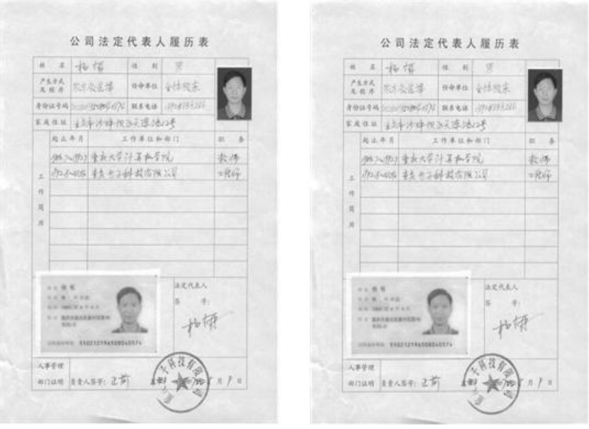
$$RI = \frac{R_{SRC} - R}{R}$$

Where R and R_{SRC} were the compression ratios produced by individual page compression techniques and by SRCs. From table 1, we know that SRC such as MMD and CTD could compress the images more effective than individual page compression techniques. Moreover, the CTD's compression ratio was higher than MMD's.

TABLE II. COMPRESSION PERFORMANCES BASED ON CTD

Compared. method	Average size (Byte)	Average Compr. ratio	Compr. Ratio Impr. (%)	Average PSNR (dB)
No Compared	959274	-	-	-
JPEG2000	95927	10.00	-	33.9115
MMD +JPEG2000	84073	11.410	14.10	33.9115
CTD +JPEG2000	82200	11.670	16.70	33.9115
JPEG2000	25600	37.47	-	29.5835
MMD +JPEG2000	22786	42.098	12.35	29.5835
CTD +JPEG2000	22419	42.787	14.19	29.5835

When $PSNR = 33.9115dB$, the solution of CTD was illustrated in figure 7. If the images were amplified for detail observation, you could not find any visual difference by human eyes. When the $PSNR$ decreases to 29.5835dB, we can observe the obvious distortion in these two images. $PSNR = 29.5835$ dB corresponding to an actual application requirements, which is the size of every image compressed, is less than 25KB.


公司法定代表人履历表
公司法定代表人履历表

(a) By JPEG2000 (b) By CTD+JPEG2000

Figure 7. Compression solution based on CTD (PSNR=33.9115Bd)

VI. CONCLUSION

Compression is based on the elimination of data redundancies. Sets of similar images contain a special type of redundancy. The "set redundancy" which is defined as the common information can be found in more than one image of the set. Set redundancy compression utilized not only the statistical information in individual page, but also the information in page-page. The more similarity the images have in the set, the higher compression ratio the SRC produces. CTD build the similar image set through template which is associated with the attributes of document image. CTD is essentially the improvement of MMD. It could be apply to some DIISs in industrial and commercial enterprise

It should be noted that, the SRC methods impose high similarity in the whole set of images. A preprocessing phase can be done to cluster similar images before launching the compression operation. SRC methods can also be tested on many other application areas. Satellite image databases, for example, often contain sets of images with the same geographical areas, similar weather and lighting conditions. They necessarily contain page-page redundancy.

Which compression method is selected depends on the special application. If legal or historical documents are existed in the database, the compression should be lossless. If the application is based on internet transmission or information retrieval in MIS, the compression could be loss. CTD could be regarded as a preprocessing procedure before the compression of individual page coding. It could be either lossless or loss. In the DIIS, storage size should be the goal of compression, besides some other factors such as image retrieval and image transformation should be the goal also.

ACKNOWLEDGMENTS

This work is partially supported by the Applying Basic Research Program of Chongqing Education Committee (Grant No. KJ100623) to Ping Yu, and the Doctor Foundation of Chongqing Normal University (Grant No. 10XLB006) to You Yang.

REFERENCES

[1] You Y. Digitalization of Industrial and Commercial Enterprise Documents. Journal of Chongqing Normal University(Natural Science Edition), 2004, Vol.21, No.2, pp.31-34. In Chinese.

[2] Karadimitriou K. Set redundancy, the enhanced compression model, andmethods for compressing sets of similar images. Ph.D.thesis, Department of Computer Science, Louisiana State University, Baton Rouge, La, USA, August 1996.

[3] Karadimitriou K., Tyler J. M. The min-max differential method for large-scale storage and compression of medical images. Proceedings of of Annual Molecular Biology and Biotechnology Conference, Baton Rouge, La, USA, 1996.

[4] Karadimitriou K., Tyler J. M. Min-Max Compression Methods for medical image databases. ACM SIGMOD Record, 1997. Vol.26, Issue 1, pp.47-52.

[5] Karadimitriou K., Tyler J. M. The centroid method for compressing sets of similar images. Pattern Recognition Letters 19, 1998, pp.585-593.

[6] Jiann-Der Lee, Shu-Yen Wan, Cherng-Min Ma. et al. Compression Sets of Similar Images Using Hybrid Compression Model. Proceedings of 2002 IEEE International Conference on Multimedia and Expo(ICME'02), Lausanne, Switzerland. Aug. 26-29, 2002, Issue 1, pp.617-620.

[7] Ait-Auodia S., Gabis A. A comparison of set redundancy compression techniques. EURASIP Journal on Applied Signal Processing, 2006, pp.1-13.

[8] Nielsen C., Xiaobo L. MST for lossy compression of image sets. Proceedings of the Data Compression Conference (DCC'06). Vienna, Austria. March 28-30, 2006, pp.463.

[9] Haneda E., Bouman C. A. Text Segmentation for MRC Documnet Compression. IEEE Transactions on Image Processing. Vol.20, Issue 6, 2011, pp.1611-1626.

[10] Zaghetto A., de Queiroz R. L. Improved layer processing for MRC compression of scanned documents. 16th IEEE International Conference on Image Processing (ICIP), Nov. 7-10, 2009. Cario, Egypt. pp.1993-1996.

[11] Imura H., Tanaka Y. Compression and String Matching Method for Printed Document Images. ICDAR '09. 10th International Conference on Document Analysis and Recognition. July 26-29, 2009. Catalonia, Spain. pp.291-295.

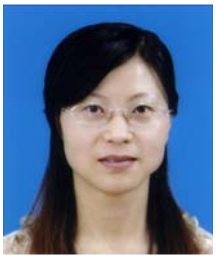
[12] You Y. A noval method to compress document image using template. Computer Science. 2008, Vol. 35, No.6, pp. 265-267. In Chinese.

[13] Salah Ameer, Otman Basir. Image compression using plane fitting with inter-block prediction. Image and Vision Computing, Volume 27, Issue 4, 3 March 2009, pp.385-39.

[14] Tanzeem Muzaffar, Tae-Sun Choi. Linked significant tree wavelet-based image compression. Signal Processing, Volume 88, Issue 10, October 2008, pp.2554-2563.

[15] Muruganandham A., Wahida Banu R. S. D. Adaptive fractal image compression using PSO. Procedia Computer Science, Volume 2, 2010, pp.338-344.

[16] Ashutosh Gupta, Suneeta Agarwal. A fast dynamic compression scheme for natural language texts. Computers & Mathematics with Applications, Volume 60, Issue 12, December 2010, pp.3139-3151.



Ping Yu received the master degree in system theory in 2005 in Chongqing Normal University, Chongqing, China. She is currently an Associate Professor of computer and information science. Her main research interests are data mining and image processing.

You Yang received the doctor degree in computer science in 2010 in Beihang University, Beijing, China. He is currently an Associate Professor of computer and information science. His main research interests are digital image processing and pattern recognition.