2006 IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan

TEXT EXTRACTION ON CHINESE PAINTINGS

Shwu-Huey Yen, Wen-Tsung Tsai, Chiu-Hsing Liu, Hwei-Jen Lin, Chia-Jen Wang

***Abstract*—This paper presents a scheme to extract inscriptions from a traditional Chinese painting such that the inscriptions and the painting can be enjoyed or studied separately. A two phases morphological operation is used to remove most content of a painting (i.e. background) which makes inscriptions to become the principal object in the remaining image. Since inscriptions are written vertically, we use the alignment property to construct the center point map and use it to locate character lines. Character block is formed by clustering adjacent character lines. The proposed algorithm has been executed on a set of Chinese paintings and proved its efficacy.**

I. INTRODUCTION

Starting around 4000 B.C. traditional Chinese painting has developed continuously over a period of more than six thousand years. Its growth has inevitably reflected the changes of time and social conditions, hence it becomes one of the most important and precious historical monument. One of the distinctive characteristics of Chinese painting is the use of inscriptions in poetry of calligraphy and of special seals as part of the painting itself. Its significance lies in its ability to express the theme and artistic conception of the painting more clearly and deeply while, at the same time, giving great insight into the artist's individuality, emotions and views on art and life [1]. Thus while taking pleasure in the content of Chinese painting, it is also important to enjoy the inscription and seals at the same time. Thanks to the modern developments of multimedia technology and digital archives programs, progressively more and more Chinese paintings and other cultural treasures are digitized.

Consequently, not only they can be preserved and utilized in the digital era but also they can improve literacy, creativity and life quality of mankind. In this paper, a novel scheme is proposed to extract the inscription and seals from a Chinese painting such that people can enjoy the content of painting and the inscription separately, as illustrated in Fig. 1. Since inscriptions on a Chinese painting are similar to texts on an image, from here on we will refer it as texts.

S. H. Yen is with the PRIA Lab., Department of Computer Science and Information Engineering, Tamkang University, Tamsui, Taipei, Taiwan (Tel: 886-2-26215656, ext. 3305; fax: 886-2-26209749; e-mail:shyen@cs.tku.edu.tw).

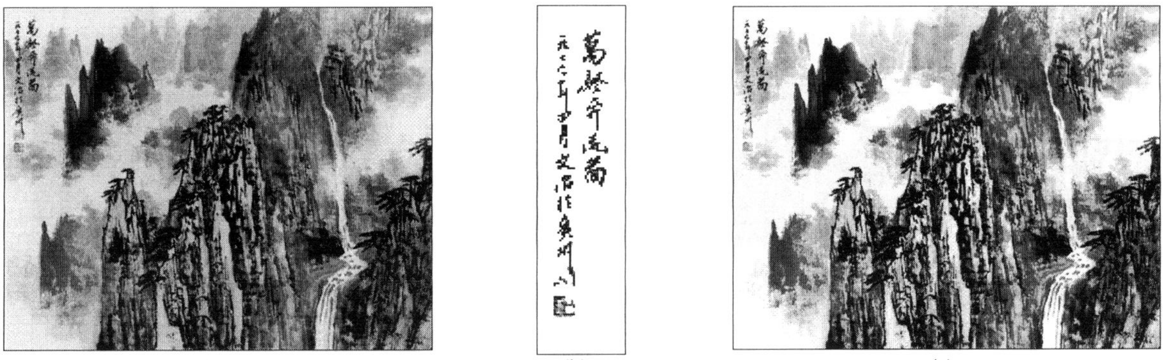
C. H. Liu was with the PRIA Lab., Department of CS & IE, Tamkang University.e-mail:692192064@s92.tku.edu.tw

W. T. Tsai, H. J. Lin, and C. J. Wang are also with the PRIA Lab., Department of CS & IE, Tamkang University (e-mail: 490191631@s90.tku.edu.tw, {hjlin,121718}@mail.tku.edu.tw).

The text extraction task on a Chinese painting is a challenging subject. Firstly, inscriptions were brush-written by scholar painters, characters usually are not in identical sizes (as in Fig. 1 (b)) and not necessarily in perfect alignment either. Secondly, due to long period of time of original artwork and reproduction technique, it is common that the acquired painting images suffer the uneven illumination problem. In addition, inscriptions may be overwritten on painting content (as in Fig. 1(a)) causing characters and background mixing problem.

Text extractions in gray-tone images usually refers to binarization problems. Local or adaptive binarization typically is applied to extract text when it is embedded on complex background [2]– [4]. Nevertheless, text mixed with background still is a difficult problem due to many different intricacies of images. Lately many text-extraction researches are extended to color images [5]-[10]. A common approach is to reduce the number of colors from 24-bit true colors to principal colors, and text regions are identified from each principal color plain followed by text regions merging [8]-[10]. These methods usually require characters are in solid color and not too small, and background can not be toocomplicated. Yen *et al.* [11] proposed a method to extract text printed on dated color postcards. Texts on postcards suffer problems of text-size too small, uneven illumination, poor reproduction quality, etc. Although some difficulties are similar as in this research, but the sizes of text and interline spacing are more uniform because texts on postcards are printed.

There is no proposed algorithms can solve all the problems of text extraction. The complexity of text extraction, to name a few, lies in the cases that texts mixed with backgrounds, texts with varying sizes, color, font, and uneven illumination, etc. Texts extraction in Chinese paintings also faces same difficulties except the text color. Since inscriptions of Chinese paintings are always brush-written by black ink, thus intensity is a good feature for locating texts. In our study, we use gray-tone images of paintings for text extraction. The paper is organized as follows. Section II describes the overall proposed algorithm. Experiments and discussions are in Section III and finally the conclusion is given.



(a) (b) (c)

Fig. 1. (a) The original Chinese painting, (b) the extracted text, and (c) after enhancing intensity of background pixels.

II. THE ALGORITHM

There are many different styles of traditional Chinese paintings. In here, we are only interested in those paintings such that the painting, referring as background, is primary and inscription, referring as text or foreground, is secondary in the whole painting. An Inscription on a painting is written by Chinese calligraphy in ink, it could include a piece of poem, or simply the name of the painter, or both. Generally speaking, the area of inscriptions on a painting is not more than 5% of total area of a painting for most of cases. Since texts are darker than the background, we first make those background pixels brighter to enhance the difference between text and background. The image is first converted from RGB into HSI, then for every pixel,

$$\textit{if} \quad (I>115)\ or\ (I\geq 90\ and\ S>0.125) \quad then \quad I\leftarrow(1.25)*I \qquad (1)$$

where I, S are intensity, saturation values of pixels on HSI model.

According to our observation, if a pixel is bright enough, i.e. $I > 115$, or not so bright but it is saturated, i.e. $I \geq 90$ and $S > 0.125$, then it is very possible to be a background pixel. The result of applying Eq. (1) is as shown in Fig. 1(c). Details are described in the following.

A. Background Removal

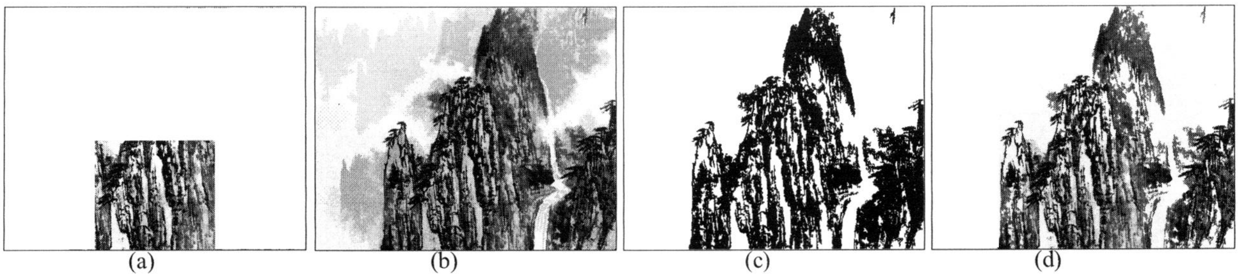
Two phases morphological reconstruction is used to reconstruct most of background. Then the goal of removing irrelevant background can be achieved by the difference of the original image and the rebuilt image. Since the background is brighter than the text, thus background is reconstructed by erosion. Given two gray-scale images, a marker image $\boldsymbol{M}$ and a reference image $\boldsymbol{R}$ of the same size with intensity of every pixel on $\boldsymbol{M}$ is not less than the intensity of corresponding pixel on $\boldsymbol{R}$, to reconstruct $\boldsymbol{R}$ from $\boldsymbol{M}$ by erosion is repeating a process of erosion to $\boldsymbol{M}$ followed by a point wise maximization with $\boldsymbol{R}$ until $\boldsymbol{M}$ is stable [12].

In most of traditional Chinese paintings, inscriptions are hardly connecting to four sides neither on the lower center region of a painting. Thus the reference image $\boldsymbol{R}$ is the gray scale image from the enhanced image, and the marker image $\boldsymbol{M}_1$ is a white image with pixels of four sides and the region, located in the lower 55% to the bottom of the height and the center 40% of the width as shown in Fig. 2 (a), to have the same gray values as in $\boldsymbol{R}$. The details are given below.

Phase 1: let I_1 be the reconstructed image of reference $\boldsymbol{R}$ and marker $\boldsymbol{M}_1$. Binarize I_1 with the threshold t*0.75 where t is the threshold obtained from image I_1 by Otsu's method [13]. Image I_2 is binarized mage I_1 with small blocks removed. A block is small if number of pixels within such block is not more than 50. Fig. 2 (a), (b), (c) show images of marker $\boldsymbol{M}_1$, I_1, and I_2 respectively. As can be seen in I_1, those large and dark blocks are primary elements of the painting, or primary elements of the background. To further remove background, we add these primary elements into the marker and redo the morphological reconstruction. In obtaining I_2, threshold is lowered and small blocks are removed in order to exclude those text mixed with background.

Phase 2: Define the second marker image $\boldsymbol{M}_2$ such that those pixels that in $\boldsymbol{M}_1$ or in I_2 have the same gray values as in $\boldsymbol{R}$. The marker image $\boldsymbol{M}_2$ is showed in Fig. 2(d). Redo the reconstruction by erosion to reference image $\boldsymbol{R}$ and marker image $\boldsymbol{M}_2$.

Let D be the difference of the reference image $\boldsymbol{R}$ and the reconstructed image from $\boldsymbol{M}_2$ and $\boldsymbol{R}$. D will be binarized and the resulted black objects are the potential text. Since most of background has been removed, now D has most of white background as seen in Fig. 3(a). In order to avoid the case that background mixed with text is mistaken as text, the original threshold value t_1 is modified according to the ratio of potential text area. The details of binarization process for D is described as follows.



(a) (b) (c) (d)

Fig. 2 (a) The marker $\boldsymbol{M}_1$, (b) I_1, the reconstructed image, (c) I_2, binarized I_1 with small blocks removed, consists of primary elements of the painting,

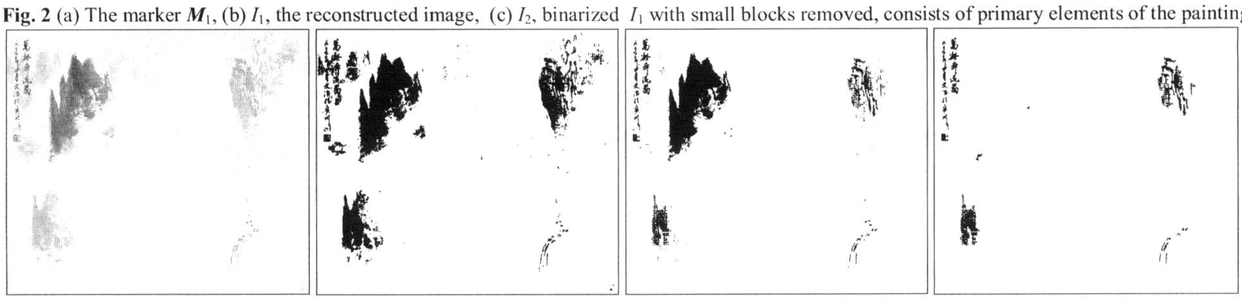


Fig. 3 (a) The difference image D, (b) binarize D using threshold t_1, (c) D_1, binarize D using threshold t_2, (d) D_2, noise removal on (c).

- Find the threshold t_1 by Otsu's method [13] for D, and evaluate the ratio of the sizes between potential text (black pixels) and background (white pixels) if D is binarized using t_1. Define a modified threshold value t_2 as

$$t_2 = t_1 \times (1 - \alpha \cdot r^2), \tag{2}$$

where r = min(ratio, 0.111), and $\alpha \geq 10$ is a constant, 11 is used in this paper.

- Let D_1 be the image obtained from the binarized D using the threshold t_2 followed by a size of 2 × 1 opening to connect broken pieces caused by binarization. As mentioned in the beginning, the area of text should not more than 5% of total area of the painting. To be safe, we allow text area to be as much as 10%, i.e. ratio = (10%)/(90%) = 0.111. If it is more than 10% then very possibly it has taken some background as text. By lowering the threshold the background mixing with text can be classified correctly. As in Fig. 3(b) and (c) are the results of D binarized using $t_1$and t_2 respectively, and the area on upper left corner shows the effect of the threshold modification.
- A simple noise removal is applied next. Those small connected components, number of pixels ≤ 10, will be removed. In addition, those large connected components, height ≥ 1/4(height of the painting) or width ≥ 1/8(width of the painting), are also removed. The result, D_2, is shown on Fig. 3(d).

B. Text Extraction

Due to vertical line up characteristic of inscriptions, using one point to represent one character, points are roughly in vertical alignment. From this fact, the vertical line with maximum number of points on it will be a text line. This line is used for the fundamental of text lines, and the height of characters can be estimated from horizontal projection. With information of location and size of characters, text block can be located by extending and merging more text lines from this fundamental line.

1) Locate text lines: Ideally one character corresponds to exactly one CC, but in fact there are cases that are not what we have expected. For example, one character breaks into pieces especially in side-by-side position, or many characters into one especially in vertical position. The former, an opening operation with a horizontal SE can help, as the latter, we define the crossing count, N, to estimate number of characters in a given CC. For any given CC, let N_1, N_2, N_3 be the number of encounters from black to white and white to black in vertical direction at 1/4, 1/2, and 3/4 width of such CC. The crossing count N is the maximum of N_1, N_2, N_3 as shown in Fig. 4(a). Since numbers of strokes in Chinese characters vary a lot, we assume approximately there are more than 2 horizontal strokes in one character, thus if CC consists of one character then we require that N > 4. The detailed steps are described below.

(1). Apply a 3 × 1opening operation to D_2.

(2). Sort all CCs according to widths and define Char_W to be the median of all widths.

(3). Center Points Map: use points located in the center of a CC to represent potential characters as shown in Fig. 4(b), the center point map of Fig. 3(d). For every CC with N, h and w as its crossing count, height, and width respectively, do the following:

- if w > 2.5* Char_W or w < 1/4* Char_W, then it is considered as a noise and removed; otherwise
- if (h/w) > 2 and N > 4*(h/w), then this CC will be represented by $\lfloor h/w \rfloor$+1 points located evenly in the vertically center of the CC; else there is only one point in the center of this CC.

(4). For every vertical line located at y-coordinate as 4k-1, k = 1, 2, 3, …, count the total number of points on vertical lines positioned at y = (4k-1), (4k-1)±1, (4k-1)±2. Define Line_1 to be the line with maximum number of points as in Fig. 4(c).

If Line_1 is located at y = 4j-1, then project all points on vertical lines positioned at y = (4j-1)±1, (4j-1)±2 to Line_1 and find out vertical distance for any two adjacent points Line_1. Sort all the vertical distances and let Char_H be the median.

(5). Obtain the image of potential text lines: Do erosion with SE size to be 1.5·(Char_W) × 3·(Char_H), followed by a dilation with size of (Char_W) × 3·(Char_H) to the center points map. By this way, isolated points, away from other points by at least 3·(Char_H) vertically, are removed. Further noise removal is accomplished by a dilation of 3 × 3. Fig. 4(d) shows the result.

2) Locate text block T: In order to merge lines into text block, an erosion is first applied to the image of text lines with size of Char_W × Char_H. Do the following with respect to the Block_1, the largest CC located on the Line_1:

(1). Obtain the text block T:

- Extend the CC by Char_H and 2*Char_W to its top-bottom and left-right borders respectively. Merge blocks into one if it touches any others.
- To include Chinese poem written (top- below) style, consider regions right on top of the block and below the block, if there is any block intersects with the region and its crossing count lies in a range of 0.8 to 1.2 of the crossing count of Block_1, it will be included in the block.
- Text block is obtained by further extending the block by Char_H and Char_W to its top-bottom and left-right borders respectively. The result is shown in Fig. 4(e).

(2). Extract the text: Correspond the text block T to enhanced painting image, Fig. 1(c), name this image F.

- Use 4 border sides of F as the marker image and F as the reference image, do the reconstruction by erosion and followed by Otsu's binarization. A simple noise removal, if the number of pixels is less than 3, is applied. The result is as in Fig. 4 (f).
- If any sides of the marker image had used previously in $\boldsymbol{M_1}$ of phase 1 reconstruction (cf. *A* of Section II), a reconstruction on F will do again with marker image without those sides. That is to locate characters with strokes connecting to border side if there is any. The final result is the combine of the two.
- For the text extracted, a simple sharpening is applied on R, G, B values to get a better result as in Fig. 1(b).

In some rare cases, characters may have strokes connecting to border of an image. Consequently, characters will be treated as background and removed. As in Fig.5, the original image (a) has "ju" (珠) connecting to border, as a result, the extracted text in (c) does not include the character "ju". The second part of step (2) is to recover missing characters. As illustrated in Fig.5, (b) is the text block T, the mask image ***M*** is the gray-toned of T. Recall that the top and right borders of T had been used for previous reconstruction. Thus, the marker image ***R*** is a white image of the same size with exceptions on points of the bottom and left borders. Those points have the same values as in ***M***. (d) is the reconstruction result, and (e) is the difference result from ***M*** and (d). (e) is the extracted text after sharpening.

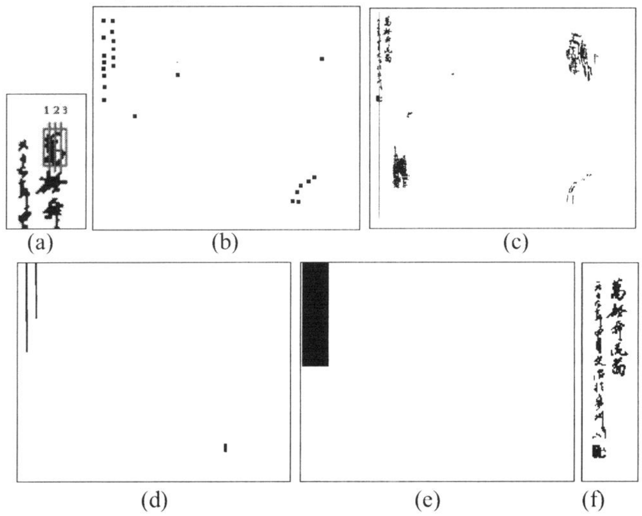


Fig. 4 The process of character block. (a) the crossing count N= max{N_1, N_2, N_3}, (b) the center point map (each point is represented by a 8x8 block here), (c) Line_1 is indicated, (d) potential text lines, (e) the text block, T, and (f) the extracted text in black and white.

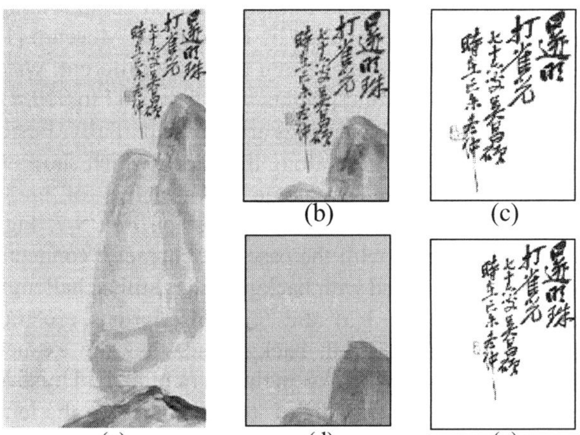
III. EXPERIMENT RESULTS AND DICUSSION

The proposed method is implemented on sample images taken from the book published by Kuboso Art Museum [14]. The dimensions of these painting are quite different, widths vary from 201 to 514 and heights vary from 267 to 919. Due to the large sizes some paintings are shown partially. Besides the example of Fig. 1 seen along the paper, which shows the proposed method can overcome the problem of uneven illumination, we exhibit more examples on Fig. 5-7. Fig. 5 explains how to deal with the case that character connecting to borders. Text mixed with background is still a challenging problem. As in Fig. 1(a) and 6(a), inscriptions are much darker than the underneath background, thus the extracted results are very satisfying except there is a plum bud mistaken as characters in the middle of the sentence on the left of 6(b). As in Fig.6(c), inscriptions are tangled with vines and leaves on the painting. The second character from top "fang" (妨) connecting to a leaf on the right, thus it is mistaken as background. The ninth and tenth characters "you" "yu" (酉, 雨) and the last four characters are extracted nicely without the influence of underneath vines.

Although most of paintings have only one text block, in some cases they may have two text blocks located in the left and right margins. Consider the corresponding gray toned image of D_2 (cf. *A* of Section II), define ***s*** as the vertical line coinciding with the right (or left) border of the text block T if such block is on the left (or right) margin of the painting. The painting is divided into two parts by ***s*** and let **I** be the one where T is not in. To complete the task of text blocks detection, **I** will be further examined for possible text blocks. For image **I**, we first remove background via reconstruction from the side ***s*** only instead of two phases. And follow the steps described in *A*, *B* in Section II to get potential character lines. For Line_1 obtained in **I**, the condition that the number of transitions **N** to be more than $4 \cdot (h/w)$ must be checked before merging character lines. If the condition is not satisfied then there is no text block in **I**. As in Fig. 7, the poem on the both sides, date and the name of the painter on the left are all extracted successfully.

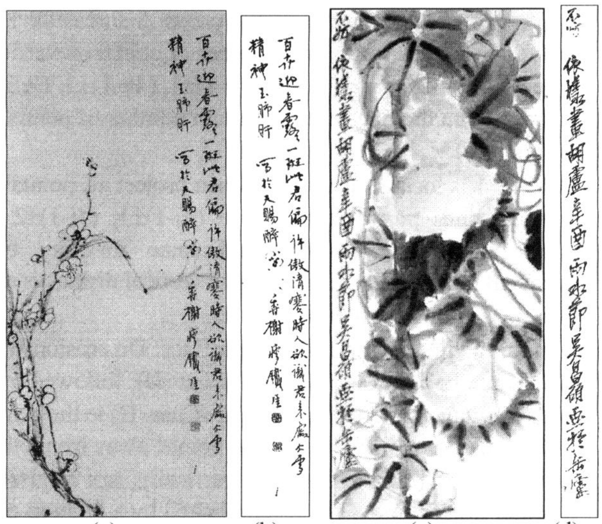
IV. CONCLUSION

A complete study on extracting inscriptions in traditional Chinese paintings is presented in this paper. Two phases morphological reconstruction and the center point map due to vertical alignment characteristic of Chinese characters are main contributions of our work. The proposed method works well on the situations of text mixed with background and uneven illumination. It is also robust to the font size of the characters. In the future, the color information will be studied in order to distinguish texts and background more precisely. The topic of inpainting after inscriptions extracted is also on the progress. To recover the original painting without distraction of added inscriptions other than the original painter is an interesting and meaningful issue.



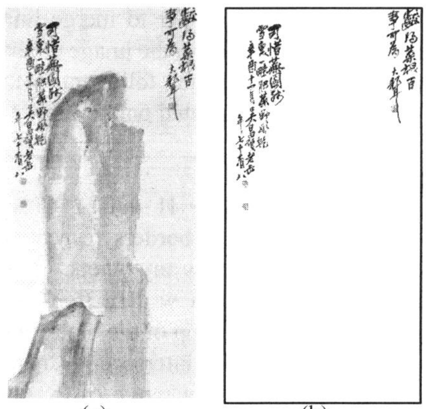
(a) (b) (c) (d) (e)

Fig. 5 (a) original image; (b) text block, its gray-toned image is used as the mask image ***M*** ; (c) text extracted before post-processing; (d) the reconstructed background and (e) is the final text extracted.



(a) (b) (c) (d)

Fig. 6 (a), (c) original images and (b), (d) the extracted text.



(a) (b)

Fig. 7 (a) original image and (b) the extracted text.

REFERENCES

[1] Asia-Art.Net, Chinese Brush Painting, http://www.asia-art.net/chinese_tech_brush.html

[2] P.K. SAHOO, S. SOLTANI, AND A. K. C. WONG, "A SURVEY OF THRESHOLDING TECHNIQUES," *COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING* 41, 1988, PP. 233-260.

[3] J. Suavely, T. Seppanen, S. Happakoski, and M. Pietikainen, "Adaptive Document Binarization," *Document Analysis and Recognition*, Proceedings of Fourth International conference, Vol.1, 1997, p. 147-152.

[4] Sittisak Rodtook, Yuttapong Rangsanseri, "Adaptive Thresholding of Document Images Based on Laplacian Sign," *International Conference on Information Technology: Coding and Computing*, April 02-04, 2001, pp.501-505

[5] Kongqiao Wang, J.A. Kangas, Wenwen Li, "Character segmentation of color images from digital camera," *IEEE Document Analysis and Recognition*, 2001. Proceedings. Sixth International Conference on, 10-13 Sept. 2001, pp. 210-214.

[6] Chun-Ming Tsai and His-Jian Lee, "Binarization of Color Document Images via Luminance and Saturation Color Features," *IEEE Transactions on Image Processing*, Vol. 11, No. 4, April 2002, pp. 434-451.

[7] E. K. Wang, Minya Chen, "A new robust algorithm for video text extraction," *Pattern Recognition*, Volume: 36, Issue: 6, June, 2003, pp.1397-1406

[8] C. Strouthopoulos, N. Papamarkos, A. E. Atsalakis, "Text Extraction in

Complex Color Documents," *Pattern Recognition* Volume: 35, Issue: 8, August, 2002, pp. 1743-1758.
[9] Wei-Yuan Chen, Shu-Yuan Chen, "Adaptive Page Segmentation for Color Technical Journals' Cover Images," *Image and Vision Computing* 16, 1998, pp. 855-877.
[10] Kongqiao Wang, Jari A. Kangas, "Character location in scene images from digital camera," *Pattern Recognition*, Volume: 36, Issue: 10, October, 2003 pp.2287-2299.
[11] S. H. Yen, M. F. Chen, H. J. Lin, C. J. Wang, C. H. Liu, "The extraction of characters on dated color postcards," *ICME 2004*, pp. 1415-1418.
[12] P. Soille, "Morphological Image Analysis: Principles and Applications," Springer, Alemanha, 1999.
[13] N. A. Otsu, "Threshold selection method from gray-level histograms," *IEEE Trans Systems, Man and Cybernetics*, Vol. 9, No. 1, 1979, pp. 377-393.
[14] Kindai Hiyakunien Chugoku Kaiga: Teiseido (One Hundred Years of Chinese Modern Painting from the Teiseido collection), Kuboso Art Museum, Izumi, Osaka, Japan, 2000.