Character Image Extraction from a Scene Image Using Color Information and Multiscale Analysis

Shinichiro Omachi Yuusuke Matsuda Hirotomo Aso
Graduate School of Engineering, Tohoku University
Aoba 6-6-05, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan

Abstract

In this paper, a method for extracting a character image from a scene image using color information and multiscale analysis is proposed. In scene images, it is difficult to separate a character image from the background accurately because of noises and the change in the lighting condition. In the proposed method, first the number of colors is reduced by a color clustering technique and several binary images are generated from the clustering results. Meanwhile, a character region is estimated by using the edges detected from the image. In the estimated region, a global structure of the character is obtained. Finally, the binary image that represents the character is obtained by unifying the information of the global structure and the binary images. The effect of the proposed method is evaluated with the experimental results using the ICDAR2003 datasets.

1. Introduction

Detecting and recognizing characters in a scene image accurately are still a challenging problem in the field of document analysis [6, 8]. A lot of techniques for detecting characters or strings in a scene image have been proposed so far [5, 7, 11, 12]. However, even if characters are detected, recognizing them is another difficult task. In this paper, we propose a method for extracting a clean character image from a scene image in order to recognize the character accurately. If a document image is obtained by an optical image scanner, it is not difficult to separate characters and the background. However, in the case of scene images, lighting condition, blurring and existence of many objects similar to characters make it difficult to separate characters and the background. The proposed method utilizes color information and multiscale images of the input image.

Figure 1 displays the outline of the proposed method. A scene image including a character is given. Since a character in scene images usually consists of only one color, it is useful to separate a character color and the colors of the background for extracting a character image. For this purpose, a color clustering technique [4] is applied and the colors of all the pixels are converted to one of the representative colors. Next, multiple binary images are generated by using the color information. Meanwhile, a character region is estimated with the edges detected in the image. In the estimated region, a global structure of the character is extracted by using the multiscale images [10]. Finally, the binary image that represents the character is obtained by comparing the global structure and the binary images. The effect of the proposed method is evaluated by the experimental results using the ICDAR2003 datasets.

2. Proposed method

2.1. Generating binary images

In order to eliminate noises and make the same objects (e.g., characters, background, shadows, etc) have the same RGB values, a smoothing method with a median filter, which is called the Weighted Median Filter based Anisotropic Diffusion [2], or WMFAD, is applied. The WMFAD can smooth an image without discarding edges. Then a color clustering technique [4] is applied to the image. The colors of all the pixels are projected into the L*a*b* color space in order to reduce the effect caused by the illumination, and a cube including all the pixels is determined by measuring the maximum and minimum values of L*, a* and b*. Then the cube is divided into 8000 subcubes by dividing each edge into 20 segments. The number of subcubes was determined experimentally.

Next, the number of pixels of each subcube is counted. If the number of pixels in a subcube is larger than the ones of its surrounding subcubes and it is larger than a threshold, a representative color is determined by calculating the gravity of the pixels in the subcube. Here, the threshold was determined experimentally. Considering all the subcubes that satisfy this condition, we can get a set of representative colors. Then, the color of every pixel is converted to the

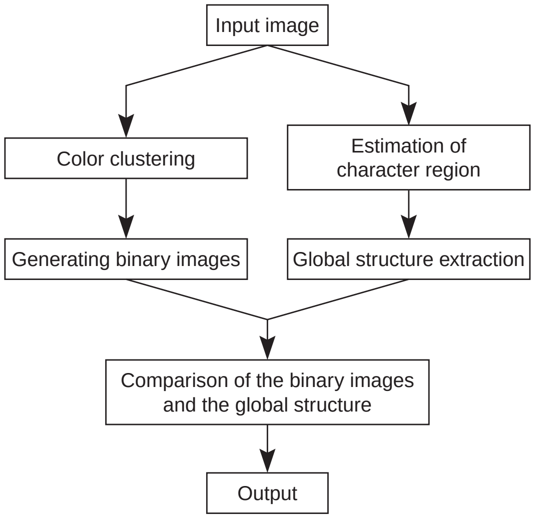


Figure 1. Outline of the proposed method.

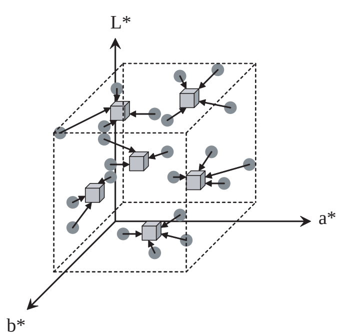


Figure 2. Color clustering.

nearest representative color. Figure 2 displays an example. The circles show the colors of the original image, and the cubes are the representative colors. Finally, the L*a*b* values of the representative colors are converted to the RGB values.

Every color in the input image is converted to one of the representative colors by the color clustering. By this procedure, the number of colors is decreased. In some cases, however, two colors have almost the same RGB values caused by the differences between RGB and L*a*b* spaces. Therefore, the number of colors in the RGB space is reduced by converting two similar representative colors into one of them. The condition of selecting similar colors was determined experimentally as follows.

- The differences of R, G and B values between the colors are less than or equal to five.

- The difference of luminance is less than or equal to ten.

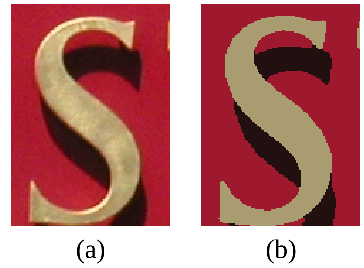


Figure 3. An example of the reduced color image.

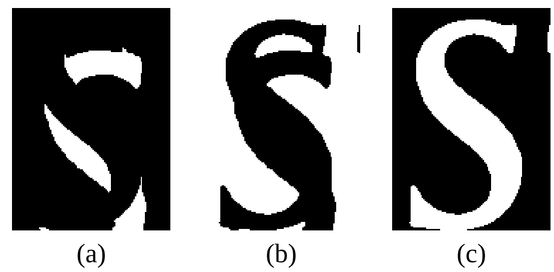


Figure 4. Binary images generated from the reduced color image.

For two colors satisfying these conditions, the numbers of pixels of the colors are counted. The color which has fewer pixels is converted to the other color. An example of the reduced color image is shown in Figure 3. Figure 3(a) is an original image and Figure 3(b) is the image that has only three colors generated from (a).

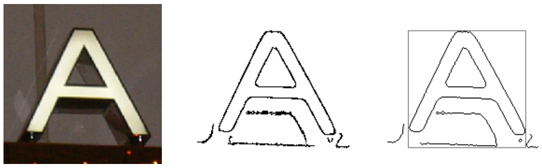
After reducing the number of colors, binary images are generated by regarding one color as the character region and the other colors as the background. Figure 4 shows examples of the binary images generated from the image of Figure 3(b). The white pixels represent the character region and the black pixels represent the background.

2.2. Global structure extraction

2.2.1 Estimation of character region

The region of a character is estimated by using edge information. The method proposed by Canny [1] is used for edge detection. First, three grayscale images are obtained by applying the Gaussian filters with variances 5, 10 and 15 to the original intensity image. Then the edge images are unified in order to get a more reliable edge image.

Figure 5 shows an example. Figure 5(a) is an original image. Figure 5(b) is the image obtained by unifying the three edge images. From this image, the longest edge is



(a) (b) (c)

Figure 5. Unification of edges.

(a) (b)

Figure 6. Reversing intensity.

detected. The edges of which lengths are shorter than the one-fifth of that of the longest edge are eliminated as noises. The minimum rectangle that includes all the remained edges is regarded as the character region (Figure 5(c)).

2.2.2 Generating grayscale images

First, the image is reversed so that the brighter pixels represent a character, if necessary. For example, the image of Figure 6(a) is reversed to the image of Figure 6(b).

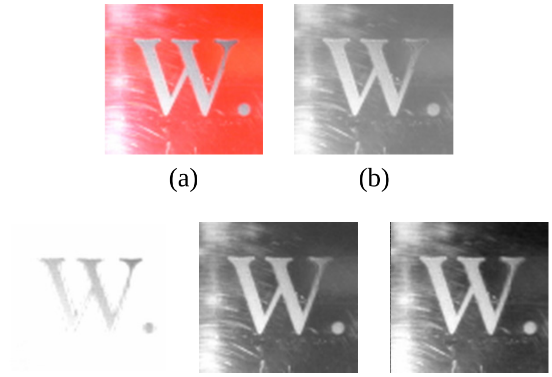
Then four grayscale images are obtained using the intensities, R-values, G-values and B-values of the original image. Figure 7(b), (c), (d) and (e) represent the grayscale images of Figure 7(a). Then the image of which difference of the average values between the character region and the background region determined by the unified edges is the maximum is selected. In the case of Figure 7, the image of (e) is selected as the grayscale image from which the global structure is obtained.

2.2.3 Ridge detection

A global structure of a character is obtained by using the multiscale images [9]. By changing the scale (variance) t of the Gaussian filter, multiscale images are generated as shown in Figure 8. The image of scale t, $L(x, y; t)$, is obtained by the convolution of the image $f(x, y)$ and the Gaussian function of variance t, $g(x, y; t)$, as:

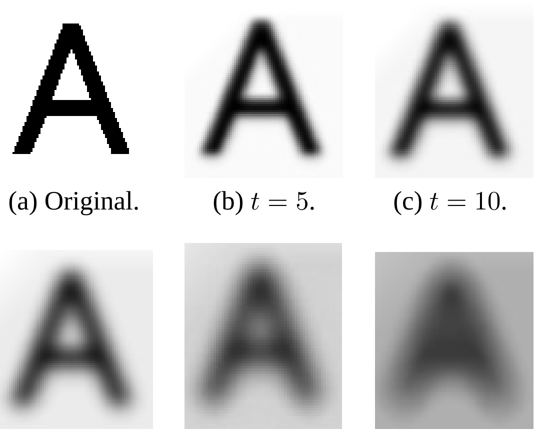
$$L(x, y; t) = g(x, y; t) * f(x, y). \tag{1}$$

Figure 9 shows the intensity surfaces [3] of the images of Figures 8(a) and (b). In Figure 9(b), it is expected that



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

Figure 7. Grayscale images.



(a) Original. (b) $t = 5$. (c) $t = 10$. (d) $t = 20$. (e) $t = 40$. (f) $t = 80$.

Figure 8. Multiscale images.

ridges of the surface represent the global structure of the character. Therefore, it is necessary to detect the ridges by selecting an appropriate scale. Let the direction of which the absolute value of the second order differential coefficient is the maximum be p, and let the direction perpendicular to p be q. The condition that the point (x, y) is on the ridges is:

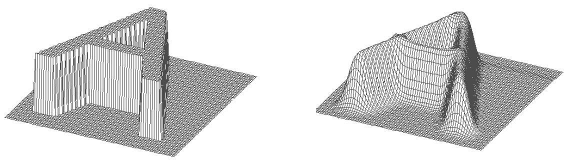
$$\frac{\partial L(x, y; t)}{\partial p} = 0, \tag{2}$$

$$\frac{\partial^2 L(x, y; t)}{\partial p^2} < 0. \tag{3}$$

The edge strength is defined as:

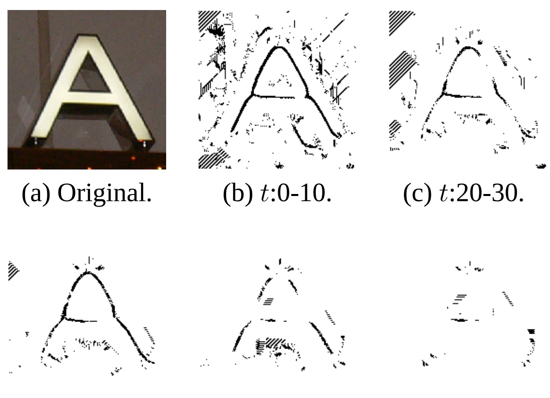
$$S(x, y; t) = \left\{ \frac{\partial^2 L(x, y; t)}{\partial p^2} - \frac{\partial^2 L(x, y; t)}{\partial q^2} \right\}^2. \tag{4}$$

The scale that maximize the edge strength is obtained by the



(a) Original. (b) $t = 5$.

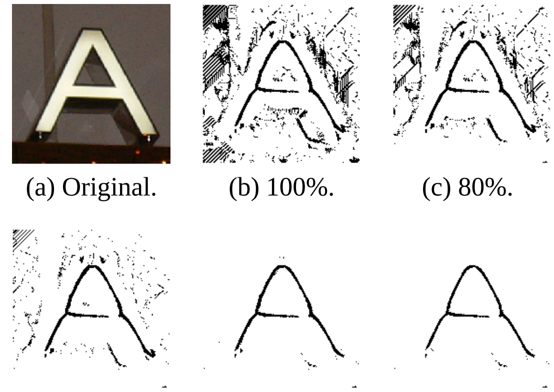
Figure 9. Intensity surfaces.



(a) Original. (b) t:0-10. (c) t:20-30.

(d) t:30-40. (e) t:30-40. (f) t:40-50.

Figure 10. Detected ridges by changing scales.



(a) Original. (b) 100%. (c) 80%.

(d) 60%. (e) 40%. (f) 20%.

Figure 11. Detected ridges by changing ridge strength.

(a) (b) (c)

Figure 12. Ridge detection.

following condition:

$$\frac{\partial S(x,y;t)}{\partial t} = 0, \tag{5}$$

$$\frac{\partial^2 S(x,y;t)}{\partial t^2} < 0. \tag{6}$$

The global structure of a character is obtained by detecting the points that satisfy Eqs.(2), (3), (5) and (6). Figure 10 displays the detected ridges by changing the scale t. Figure 11 displays the detected ridges by choosing $\alpha\%$ strong edges ($\alpha = 100 \sim 20$).

2.2.4 Extraction of character structure

In the proposed method, ridges are detected from the multiscale images by varying the scale from 1 to 30 and using 30% strong edges. Next, the edges out of the character region are removed. Figure 12 displays the results of applying the ridge detection to the image of Figure 7(e). Figure 12(a) shows the detected ridges, Figure 12(b) displays the estimated character region, and Figure 12(c) is the image that the edges outside the character region are removed.

In order to obtain a character structure from the ridge image, we have to interpolate the gaps in the image. For this purpose, ravines are detected recursively. The condition that a point (x, y) is on ravines is defined as:

$$\frac{\partial L(x,y;t)}{\partial p} = 0, \tag{7}$$

$$\frac{\partial^2 L(x,y;t)}{\partial p^2} > 0. \tag{8}$$

First an initial scale t_0 is given. Ravines are detected with this scale, and the detected ravines are added to the original image. This procedure is repeated by dividing the scale by two until the scale becomes one. Then the image is blurred. In the experiment, $t_0 = 4$. Figure 13 displays the result of recursive ravine detection. Figure 13(a) is the image after the ridge detection. Figure 13(b)$\sim$(d) are the images that ravines are detected with scale 4, 2 and 1, respectively. Figure 13(e) is the binarized image after the recursive ravine detection followed by the blurring.

Finally, in order to adjust the line width and eliminate the unevenness, thinning, blurring and binarization are applied. Figure 14 displays the whole process of the global structure extraction, and Figure 14(h) is the result.

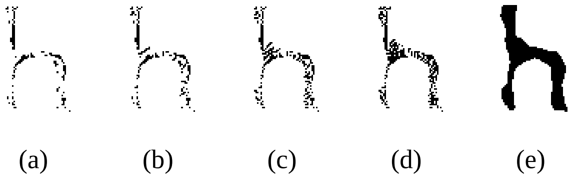


Figure 13. Recursive ravine detection.

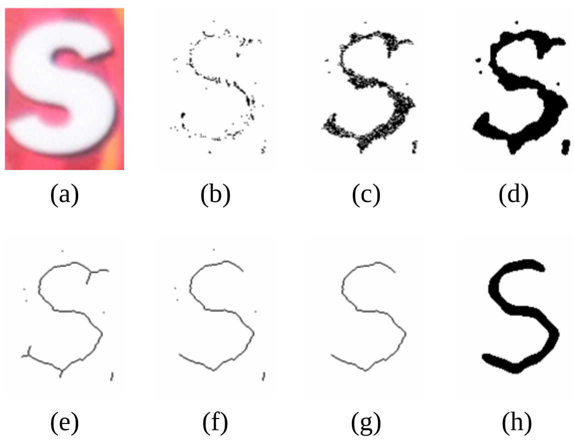


Figure 14. Character structure extraction.

2.3. Generation of the optimal binary image

From the multiple binary images obtained by the color clustering described in Section 2.1, an optimal binary image is generated. First, binary images that include the character structure more than 15% are selected. Then the selected images are overlapped. Figure 15 shows an example. Figure 15(a) is the image of which number of colors is reduced to nine. Figure 15(b) is the obtained character structure. Figures 15(c) and (d) are the selected binary images and Figures 15(e) is the final result.

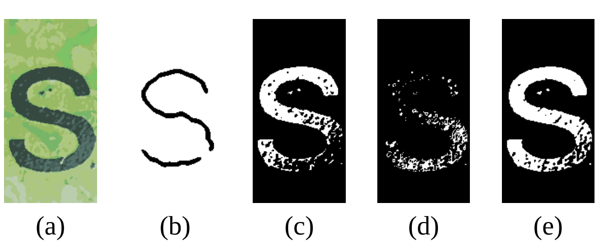


Figure 15. Optimal binary image.

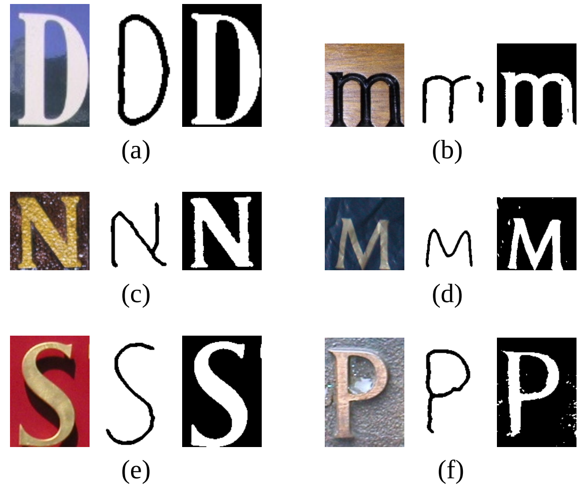


Figure 16. Results of the character image extraction (succeeded).

3. Experimental results

In order to confirm the effect of the proposed method, the TrialTrain and TrialTest datasets in the ICDAR2003 Robust OCR datasets[1] were used. Some experimental results are shown in Figure 16. In the figure, the left images are the original ones, the center images are the extracted global structures, and the right images are the results. All the character images in Figure 16 were correctly extracted by the proposed method. In these images, there is a high contrast between the character and the background, and the change in the color of the character is small. These results show the effectiveness of the proposed method for these kinds of images.

However, the proposed algorithm failed in some cases. The failed results are shown in Figure 17. In the case of Figure 17(a), a little change in the color of the character cause the failure of the color clustering. Moreover, since the image that does not have enough contrast between the character and the background, the global structure extraction also failed. Figure 17(b) is an image that has heavy change in the lighting condition, and appropriate binary images were not obtained. In these cases, a part of the character was not extracted.

From these results, it is said that a major drawback of the proposed method is the assumption that a character consists of only one color. In order to remedy this drawback, a color clustering method that works well even if a character consists of multiple colors should be developed.

[1] http://algoval.essex.ac.uk/icdar/Datasets.html

(a) (b)

Figure 17. Results of the character image extraction (failed).

4. Conclusions

In this paper, we have proposed a method for extracting a character image from a scene image by using color information and multiscale images. In the proposed method, several binary images are generated by the color clustering. Meanwhile, a character region is estimated by the edges and a global structure of the character is obtained. By comparing the binary images and the global structure, an optimal binary image that represents the character is generated. The effect of the proposed method was evaluated with the ICDAR2003 datasets. Even if there was a slight change in the lighting condition, a character image was successfully extracted by the proposed method.

Future work includes the improvement of the color clustering method and the multiscale analysis. More accurate detection of a rectangle that includes a character image is another future work.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[2] M. Gabbouj, P. Haavisto, and Y. Neuvo. Recent advances in median filtering. *Communication, Control, and Signal Processing*, 2:1080–1094, 1996.

[3] R. M. Haralick, L. T. Watson, and T. J. Laffey, The topographic primal sketch. *International Journal of Robotics Research,* 2(1):50–72, 1983.

[4] H. Hase, M. Yoneda, M. Sakai, and H. Maruyama. Consideration of color segmentation to extract character areas from color document images (in Japanese). *IEICE Transactions on Information and Systems (Japanese Edition)*, J83-D-II(5):1294–1304, May 2000.

[5] S. A. R. Jafri, M. Boutin, and E. J. Delp. Automatic text area segmentation in natural images. *Proceedings of the 15th IEEE International Conference on Image Processing*, pp.3196–3199, 2008.

[6] K. Kise, S. Omachi, S. Uchida, and M. Iwamura. Recognition and understanding of characters and documents using digital Cameras (in Japanese). *Journal of the IEICE*, 89(9):836–841, September 2006.

[7] X. Li, W. Wang, S. Jiang, Q. Huang, and W. Gao. Fast and effective text detection. *Proceedings of the 15th IEEE International Conference on Image Processing*, pp.969–972, 2008.

[8] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. *International Journal on Document Analysis and Recognition*, 7(2-3):84–104, July 2005.

[9] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Computer Society Conference on Computer Vision and Pattern Recognition*, 465–467, 1996.

[10] S. Omachi, M. Inoue, and H. Aso. Structure extraction from decorated characters using multiscale images. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 23(3):315–322, March 2001.

[11] Y.-F. Pan, X. Hou, and C.-L. Liu. A robust system to detect and localize texts in natural scene images. *Proceedings of the Eighth IAPR International Workshop on Document Analysis Systems*, pp. 35–42, 2008.

[12] L. Xu, H. Nagayoshi, and H. Sako, Kanji character detection from complex real scene images based on character properties. *Proceedings of the Eighth IAPR International Workshop on Document Analysis Systems*, pp. 278–285, 2008.