MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

3—17 Book Recognition from Color Images of Book Shelves

Yukari Akiyama and Minoru Ito
Dept. of Electronic Engineering, Kogakuin University
1-24-2, Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-8677 Japan
E-mail: ito@ee.kogakuin.ac.jp

Abstract

This paper describes a new approach for book recognition, where color images of the spines of books are analyzed to identify books displayed on book shelves. First, edges of HSY images are unified to one edge image considering the color features of the spines of each book. The color information in the images is effective for segmenting images to robustly extract book boundaries. The approach may resolve the edge extraction problem that occurs in analyzing grayscale images. Then, book title character string regions are extracted using a min/max filter. The filter enables us to extract the regions while avoiding the effects of shading distribution due to non-uniform lighting and the convex surfaces of book spines. In the last step, the cumulative projection profile of each title character string region is compared with that in a database, and the title can be determined without having to recognize each character one by one. The approach presented in this paper makes it possible to visually inspect book shelves, and paves the way for the development of an automated system for managing books in bookstores, libraries, storerooms, and offices.

1 Introduction

Recently, the recognition of characters in car number plates and newspaper headlines and the recognition of degraded print characters have been studied by many researchers [1-3]. The title character string on the spine of each book displayed on book shelves should also be considered an important target for character recognition. Managing the large numbers of books in bookstores, libraries and storerooms is tedious work: it takes a long time to confirm the locations of books on book shelves and record the locations in databases. The important issue is how to easily keep track of books as they are moved in and out, removed from and returned to shelves, and rearranged. The former strategy involves attaching a label to the spine of each book, but reading so many book labels one by one is still difficult. Hence, a new strategy by which the title strings on the spines can be visually and automatically recognized without having to handle the books is strongly required.

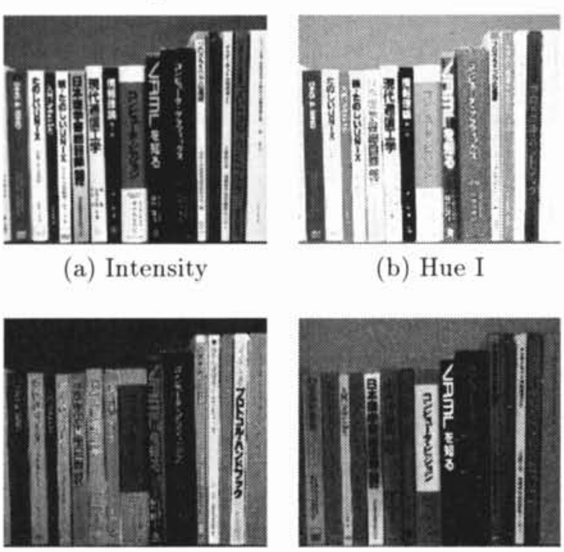
To extract book boundaries from gray-scale images of book shelves, we previously proposed a method in which the length and continuity of each edge line are examined to confirm whether the line is a true book boundary or not [4]. Tominaga et al. used Hough transformation to extract longitudinal boundary edges [5]. However, line components along title character strings could not be eliminated from all of the line candidates obtained. As a method for extracting the book title string regions from the spines of books, we proposed min/max filtering. This approach effectively avoids the shading distribution on the spine. Sawaki et al. used projection profiles for the same purpose [6, 7].

Fortunately, book boundaries are composed of straight lines and the title string of each book is lined up along the center line of the spine. We can make use of these features to extract the boundaries and strings. However, boundaries and book title strings in images are not necessarily extracted as desired, especially when two adjacent books by chance possess similar density values in an image or when there is an unnegligible shading distribution in images due to nonuniform lighting or the specularity of book surfaces. These cause false edges to appear as the boundaries of books and string regions. Therefore, there is a strong need for a sophisticated new algorithm for the robust extraction of book boundaries and character string regions.

This paper proposes a new approach for book recognition, where HSY color images of the spines of books together with the edge direction distribution are analyzed to extract book boundaries [8]. In addition, title character string extraction by using the min/max filter and book title identification by comparing the projection profile of each string with that in a database are demonstrated.

2 HSY image acquisition

The color image of books on a book shelf is acquired with the conventional film camera and the negative is read with a film scanner to obtain the RGB digital image. The image is 512×480 in size and the depth is 8 bits for each color. A large piece of pink paper is placed deep in the book shelf to avoid the unstable hue values associated with dark backgrounds. RGB values of the input image are converted to HSY values to generate H (hue) image, S (saturation) image and Y (intensity) image. This is done according to the relations. $Y = 0.3R + 0.59G + 0.11B$, $H = \tan^{-1}(C_1/C_2 - \alpha)$ and $S = \sqrt{C_1^2 + C_2^2}$ where $C_1 = R - Y$ and $C_2 = B - Y$, where α is a parameter giving the initial phase to the hue. The difference of the hue angle for each pixel color from the hue angle α is represented by $|H|$. The HSY images used as the basic images in this paper are shown in Fig. 1, where (a) is the intensity image, (b) hue image I when α is 0 degrees, (c) hue image II when α is 90 degrees, and (d) the saturation image.



(a) Intensity (b) Hue I

(c) Hue II (d) Saturation

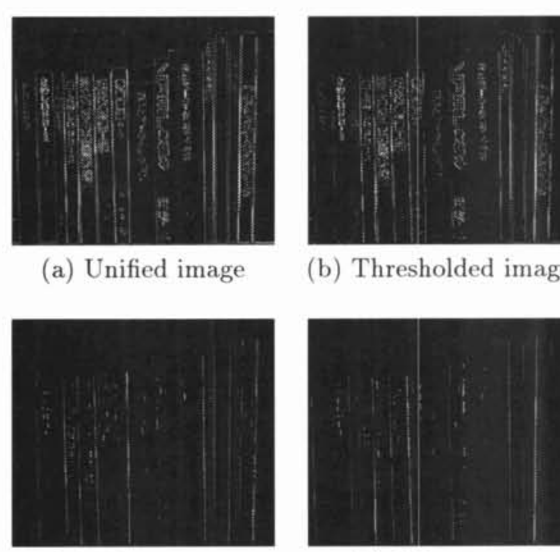
Figure 1: Basic images.

3 Extraction of book boundaries

For each of the four images in Fig. 1, the edge strength distribution is calculated using a 3×3 spatial filter and then normalized by the maximum edge strength level of each image. By using the hue and saturation image edges, two adjacent book regions having a similar gray level (intensity) can be distinguished because the hue and saturation images represent the peculiar characteristics of the color of the spine of each book. Considering this, the four edge images are unified to one edge image so that the resultant unified edge image may represent all of the true edges of book boundaries. For this purpose, edges are unified in a way that the maximum pixel value among the four images is picked up at each common coordinate (x, y) and set as the pixel value at the same coordinate of the unified image; i.e.,

$$f(x,y) = max\{f_i(x,y)|i = 1,2,3,4\} \quad (1)$$

The resultant unified edge image is shown in Fig. 2(a). Image (b) is the binary image thresholded by an empirically given level. In (a) and (b), many of the edges of title characters are stronger than those of book boundaries. If the threshold level is decreased, more book boundary edges than title character edges disappear. Consequently, it becomes difficult to distinguish the book boundaries from the characters. Images (c) and (d) show the longitudinal edges obtained by the filter for calculating the edge direction at each pixel. The directions of pixel density tilt are right-to-left and left-to-right for (c) and (d), respectively. It is clear that edges along the title characters are largely eliminated and those of the book boundaries become dominant. This is expected to make it easy to extract true book boundaries.



(a) Unified image (b) Thresholded image

(c) Edge image (d) Edge image

Figure 2: Edge images.

Hough transformation is used to detect straight lines in the edge image. Ideally, the tops of the voting level should sharply appear at specific positions in the Hough plane with correspondence to the locations and tilts of the true straight lines. Furthermore, the number of the tops should be equal to that of the true book boundaries. However, in actual experiments, the tops often appear as dull bumps because of noise, digital sampling, and dull book boundary edges in the input image. Many more tops than the true boundaries appear in the plane. To cope with this problem, bump regions are first found in the Hough plane and then the top position is estimated in each region. In addition, as book boundaries are for the most part longitudinal in the input image, only the top positions corresponding to the lines having tilts within 20 degrees from longitudinal are selected and converted to the equations of the straight lines as boundary line candidates. This conversion is called inverse Hough transformation.

Figure 3 shows voting levels at various voting positions. The horizontal axis represents voting positions sorted in order of high-to-low voting levels, and the longitudinal axis is the voting level. Voting positions over the 40th are neglected. Curve a in this figure is the result obtained through Hough transformation for the Fig. 2(b). Curves b and c are for Figs. 2(c) and (d). There is a dominant step in the levels at voting positions 15 to 20 for curve a. In curves b and c, changes in line tilts are clearly observed at positions 15 to 20. This indicates that the voting level distribution as shown in Fig. 3 helps us determine the optimal threshold of the voting level or the number of voting positions to be converted to straight line candidates in the inverse Hough transformation.

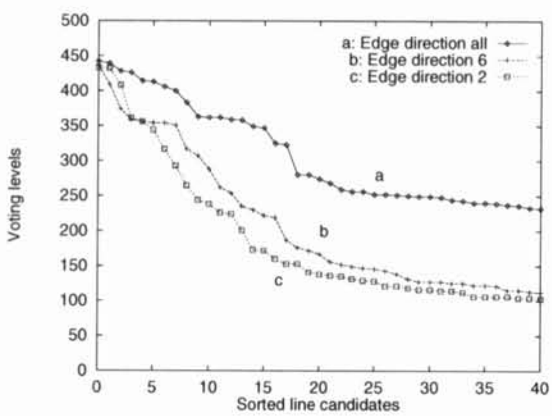


Figure 3: Voting levels for sorted straight line candidates.

Figure 4 shows the straight line candidates. The numbers of the voting positions converted are 20 for (a) and (b), and 15 for (c) and (d). The density tilt directions at the edges are right-to-left for (a) and (d), and left-to-right for (c) and (d). One should note that the straight line candidates are free from the lines along the title character strings, and the lines along the book boundaries are pretty well extracted except for one particular boundary. Some of the boundaries split into more than two adjacent lines due to noise, digital sampling, and dull edges. These adjacent lines combine to form one line. There is no clear edge line for the one boundary because one of its sides is dark brown and the other is black; the contrast for these colors is too weak to be distinguished in the intensity image. Though the boundary possesses the high contrast in the hue images, the edge at the boundary is thresholded out in edge binarization because of its low edge strength due to dull edges.

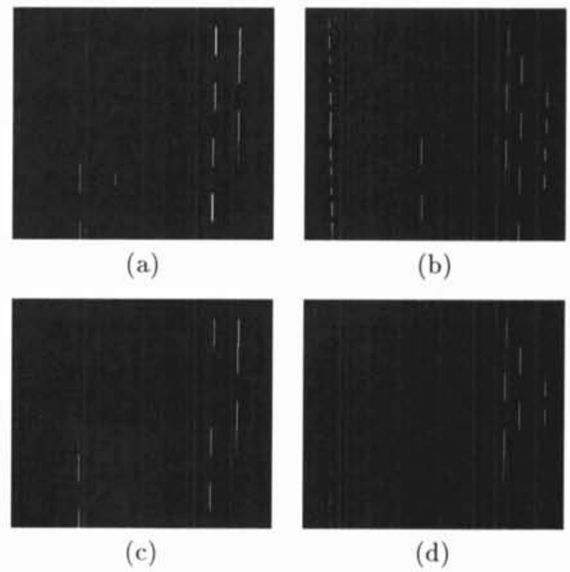


Figure 4: Detected straight lines;
(a) and (b) are for extracted 20 lines, (c) and (d) 15 lines. (a) and (c) are for one edge direction, (b) and (d) for the reverse direction.

The upper and lower boundaries of books are found as follows. The horizontal edge is first found by searching the upper part of the image along the center line of each book region surrounded by the estimated book side boundaries. Then, continuous and horizontal edges adjacent to the searched edge are counted and if the count is over half the width of each book, they are taken as the upper boundary. Meanwhile, to find the lower book boundaries, the horizontal straight line candidates are extracted from the edge image including only the horizontal direction edges by using the Hough transformation in the same manner as in the book side boundary extraction. The most dominant line is taken as the lower part of the books on the shelf.

4 Extraction of title character strings

In the former method for extracting title character strings, the binarization threshold level is set for each region of the spines of books. However, the method is not applicable when the background level of the title character string is not uniform on the spine of a book. In this section, we show that the approach based on min/max filtering deals well with background non-uniformity.

First, each book region extracted in the previous section is trimmed one by one from the input image and the trimmed images are exposed with the min/max filter. If the character string level is higher than the background level in intensity, the minimum filter is scanned on the trimmed image. If not, the maximum filter is scanned. The density histogram of each region gives explicit information about which filter should be applied for each book region. The amount of filtering required for eliminating title characters depends on the line width of the characters. For instance, when the line width is five pixels, the filtering has to be repeated three times to completely eliminate character strings. In actual experiments, the four repetitions were needed on average. Further, three repetitions were needed for extremely degraded characters whose image could not be decomposed into the line components. With the filtering, all parts of title characters are exchanged to the adjacent background levels and the remainder is the background distribution image. Importantly, the image includes almost the exact background shading distribution on each spine. Next, the input image is subtracted by the background image. The resultant image represents the shape of the title character strings.

The experimental results of string extraction are shown in Fig. 5, where (a) is an example of book regions trimmed from the input book shelf image, and (b) is the extracted title character strings. In the figure, the densities at the string and the background are reversed so that the background is always dark. After the strings are extracted for all of the book regions, the string images are unified onto one image plane as shown in Fig. 6. It is evident that the title strings are well extracted.

(a) The book region (b) The character string

Figure 5: Title character string.

Figure 6: Extracted title character strings.

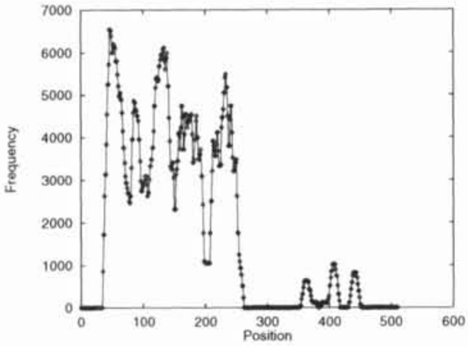
5 Book identification

Since the title characters extracted from the book-shelf images sometimes possess a low resolution, it is not necessarily easy to fully recognize the title of each book in the form of digital codes. Considering this problem, this section proposes an approach where features, such as the book dimensions and the cumulative projection profile of each title character string region, are compared with those in the database to identify each book. In preparing the database, projection profiles for images of the books to be managed are taken and normalized so that the length of the side boundary becomes equal to a certain pixel size. They are stored into the database as reference profiles together with the book dimensions.

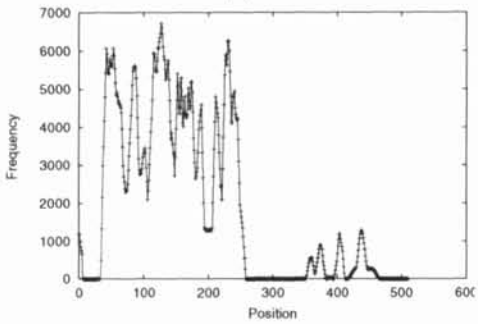
When checking books on book shelves, book boundaries and title character strings of the books are extracted as samples from the book-shelf images as mentioned in the previous section. The profile of the title character string for each book is then compared with the reference profiles for the book candidates in the database by calculating the similarities representing the correspondence between the two profiles.

Figure 7 is an example of book identification. The cumulative projection profile for the title of the sample shown in Fig. 5(a) is taken and normalized so that the pixel length of the book side boundary becomes equal to 512, and then it is averaged by an averaging filter to eliminate fine structures in the profile. The profile is shown in Fig. 7(a). We call it the sample profile. Figs. 7(b) and (c) are examples of reference profiles stored beforehand into the database. These reference profiles were normalized before being stored in the database so that the pixel length of the book side boundary is equal to 512. Here, the parameters of the first-priority region and the second priority region are given for each reference profile in the database. The first-priority region includes the clear and dominant string region of each book, such as the main title of the book, and the second-priority region includes small character strings, such as author's names.

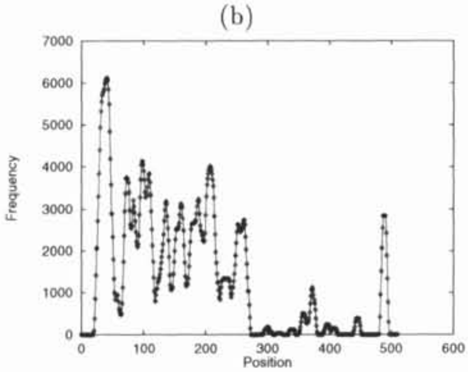
Next, the similarity between the sample profile and the reference is calculated only in the pixel region given as the first-priority region. If there are two or more reference profiles having high similarities, further comparison is made in the second-priority region to eliminate wrong candidates. To cancel the effect of small phase shifts between two profiles, the similarity is found by shifting the reference profile upward and downward. Eventually, the books in book shelves are identified from the reference books with the highest similarity.



(a)



(b)



(c)

Figure 7: The example of book indentification;
(a) the profile for the input image,
(b) and (c) the examples of reference profiles.

6 Results and Conclusion

Book recognition from images of book shelves is performed in three steps. In the first step, an input image is converted to HSY images and the edge distribution is calculated in order to unify the edge images of the HSY images. The book boundaries are then extracted using Hough transformation. The second step is the extraction of the title character string region from each book region. This enables us to extract book title strings while avoiding the effects of shading distribution due to non-uniform lighting. In the third step, the cumulative projection profile of the title string for each book is compared with the profiles for the candidates, and similarities representing the correspondence between the two profiles are calculated. Thus, the best match is found from the database to determine the title of each book.

In summary, the titles of books on shelves in camera images can be identified in the proposed approach. Our approach is available for the visual inspection of book shelves and paves the way for the development of an automated book management system for bookstores, libraries, storerooms, and offices.

References

[1] H. Onoue and M. Shiono, "A License Number Plate Extraction in a Car Image and Recognition of All Characters Containing KANJI", Transaction of IEICE, D-II, Vol.J77-D-II, No.3, pp.483–492, 1994-3 (in Japanese).

[2] M. Sawaki and N. Hagita, "Text-line Extraction and Character Recognition of Japanese Newspaper Headlines with Graphical Designs", Proc. of 13th ICPR, C73-78 (1996).

[3] K. Sawa, T. Wakabayashi, S. Tsuruoka, F. Kimura and Y. Miyake, "Low Quality String Recognition for Factory Automation", Technical Report of IEICE, PRMU97-214, 1998-01 (in Japanese).

[4] R. Iwamoto and M. Ito, "Visual Extraction of Title Character Region of Books on a Shelf", Proc. of IEICE General Conf., D-343, p.346, 1996 (in Japanese).

[5] M. Tominaga, K. Kato, K. Murakami and H. Koshimizu, "Fine Line Detection at Specified Angle by Extended Hough Transform", Proc. of IEICE General Conf., D-12-6, p.213, 1997 (in Japanese).

[6] M. Sawaki, H. Murase and N. Hagita, "Character Recognition using Automatic Template Selection based on Estimating Degradation", Technical Report of IEICE, PRMU97-187, 1997-12 (in Japanese).

[7] M. Sawaki , N. Hagita and K. Ishii, "Robust Character Recognition of Gray-Scaled with Graphical Designs and Noise", Proc. of ICDAR (1997-08).

[8] Y. Akiyama, M. Takayama, T. Nagumo and M. Ito, "Extraction of Books from Color Images" Proc. of IEICE General Conf., D-12-31, p.230, 1998-3 (in Japanese).

[9] M. Morimoto, T. Shakunaga, S. Akamatsu and Y. Suenaga, "A High-Resolution Hough Transform Using Variable Filter", Transaction of IEICE, D-II, Vol.J75-D-II, No.9, pp.1548–1556, 1992-9 (in Japanese).