Book Boundary Detection and Title Extraction for Automatic Bookshelf Inspection

Eiji TAIRA† , Seiichi UCHIDA‡ , and Hiroaki SAKOE‡

†:Graduate School of Information Science and Electrical Engineering, Kyushu University,
taira@human.is.kyushu-u.ac.jp

‡:Faculty of Information Science and Electrical Engineering, Kyushu University,
uchida@is.kyushu-u.ac.jp, sakoe@is.kyushu-u.ac.jp

Abstract In this paper, we propose a novel technique to extract individual book titles from bookshelf images for automating bookshelf inspection in libraries and bookstores. This technique is composed of two steps. In the first step, the component, such as book boundary and spine, and the local slant angle are optimally estimated at each horizontal position of a bookshelf image, and then the individual book images are isolated. For accurate estimation, a finite state automaton model is employed, each of whose states corresponds to a component of bookshelf images. In the second step, the book title is extracted from each book image isolated by the first step. A projection histogram of its edge map is utilized for this purpose.

1 Introduction

Inspection of bookshelves for managing libraries and bookstores is a time-consuming and tedious task because there are enormous books in the bookshelves and many of those books are moved frequently and repeatedly. Bar code systems are often utilized for semi-automating the bookshelf inspection. Those systems, however, are still largely based on manual tasks.

For full-automating the bookshelf inspection, several systems using bookshelf images have been studied [1]–[5]. The goal of those systems is to extract individual book information, such as title, from bookshelf images. For this goal, the following four processes are assumed in the conventional systems; (i) detection of book boundaries, (ii) segmentation of individual book images, (iii) extraction of titles, and (iv) recognition of titles.

In this paper, we propose a novel technique to extract individual book titles from bookshelf images. The proposed technique is composed of two steps. In the first step, book boundaries are detected from a bookshelf image, and individual book images are isolated. The concept of this step is to optimally estimate the component, such as boundary and spine, at each horizontal position of the bookshelf image. For accurate estimation, a finite state automaton (FSA) model is employed, which describes the structural properties of bookshelves. Roughly speaking, each state of this FSA model corresponds to a component of bookshelf image. Furthermore, for coping with non-uniformly slanted books, the local slant angle at each horizontal position is also estimated optimally.

In the second step, the book title is extracted from each book image isolated from the bookshelf image as the result of the first step. A vertical projection histogram of the edge map of the book image is utilized for this purpose.

In the conventional bookshelf image analysis systems, book boundaries are firstly detected by popular line detection techniques. For example, Hough transform is employed for detecting contiguous and straight edges as the boundaries. Unfortunately, their detection accuracy are sometimes insufficient. This is because book boundaries are often missed due to the lack of boundary edges caused by ill lighting conditions and contiguities of books of the same color. In addition, spurious boundaries are often detected due to the edges around title characters. In this paper, the boundary detection performance of the proposed technique is compared with that of the conventional technique using Hough transform.

Recently, bookshelf inspection systems using radio frequency identification (RFID) tags is developed. Those systems can be also realize full-automatic and contact-free inspection. Incorporation of bookshelf image analysis systems into the RFID inspection systems is promising because

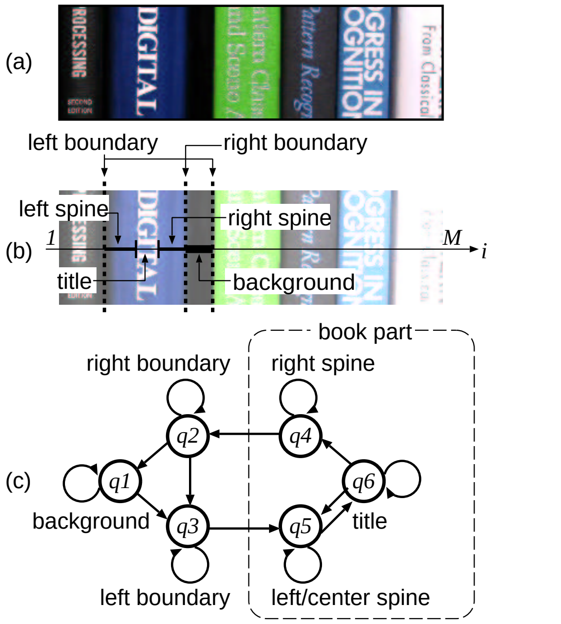

Fig. 1: (a) A bookshelf image. (b) The components of (a). (c) An FSA representation of (b).

bookshelf image analysis systems will compensate several drawbacks of the RFID tags, such as interference between tags, shield by metals, and narrow scope (due to weak wave power restricted by some law).

2 Model-based boundary detection

In this section, the first step of the proposed technique, i.e., book boundary detection (image segmentation) algorithm is presented.

2.1 Algorithm

The concept of this algorithm is to optimally estimate the component at each horizontal position of a bookshelf image such as **Fig. 1**(a). As shown in **Fig. 1**(b), bookshelf images have several components, i.e., boundary, spine, title, and bookshelf background, and the order of those components is naturally governed by some rules. The rules can be represented by an FSA model composed of 6 states ($\boldsymbol{Q} = \{q_1, \ldots, q_6\}$) as shown in **Fig. 1**(c). With this model, the estimation problem is expressed as a model-based optimization problem of the sequence $s_1, \ldots, s_i, \ldots, s_M$ where $s_i \in \boldsymbol{Q}$ denotes the state (i.e., the component) at horizontal position i.

For coping with non-uniformly slanted books such as **Fig. 2**(a), the slant angle at each horizontal position should also be estimated optimally. This estimation problem can be expressed as an optimal estimation problem of the sequence

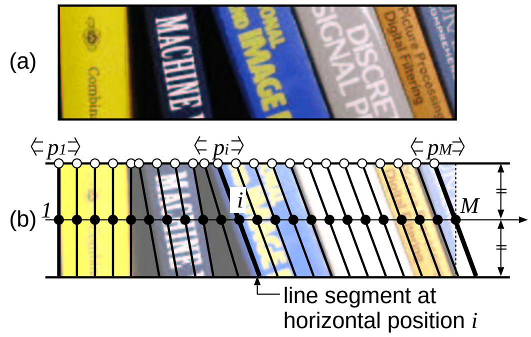

Fig. 2: (a) Non-uniformly slanted books. (b) A sequence of line segments representing the slants of books.

$p_1, \ldots, p_i, \ldots, p_M$. As shown in **Fig. 2**(b), $p_i \in [i - W, i + W]$ is the horizontal position of the top end of the line segment which passes through the center of the ith column, where W is a positive integer to specify compensable slant angles. For practical simplicity, we use p_i instead of some real-valued angle.

Those problems are formulated as the following maximization problem,

$$\textbf{maximize} \sum_{i=1}^{M} f_i(s_i, p_i), \quad \textbf{w.r.t.}\ s_i, p_i, \qquad (1)$$

subject to the transition rule of the FSA model and a constraint to limit the interval between p_i and p_{i-1},

$$p_i = \begin{cases} p_{i-1} + 1 & \textbf{if}\ s_i \in \{q_4, q_5, q_6\} \\ & \quad \textbf{or}\ s_{i-1} \in \{q_4, q_5, q_6\}, \\ p_{i-1} + \{0, 1, 2\} & \textbf{otherwise}. \end{cases} \qquad (2)$$

With this constraint, angle fluctuation is not allowed at each book part. This means the each book image has its fixed slant. On the other hand, angle fluctuation is allowed at boundaries and bookshelf background parts, while its degree is limited.

The function $f_i(s_i, p_i)$ is a criterion function to evaluate the validness of s_i and p_i at horizontal position i. The details of $f_i(s_i, p_i)$ will be described in Section 2.2.

The maximization problem (1) has the Markov property. This is because only two values s_{i-1} and p_{i-1} are necessary for determination of s_i and p_i, and the other past values (e.g., s_{i-2} and p_{i-2}) are not necessary. It is well known that the optimization problem with the Markov property can be solved efficiently using dynamic programming (DP). **Figure 3** shows a DP algorithm for the maximization problem (1). Step 7 is so-called DP recursion, and its two variables s_{i-1} and p_{i-1} are

/* Initialization: */
1 **for all** $s_1 \in \boldsymbol{Q}$ **do**
2 **for** $p_1 := 1 - W$ **to** $1 + W$ **do**
3 $g_1(s_1, p_1) := f_1(s_1, p_1)$
/* DP Recursion: */
4 **for** $i := 2$ **to** M **do**
5 **for all** $s_i \in \boldsymbol{Q}$ **do**
6 **for** $p_i := i - W$ **to** $i + W$ **do begin**
7 $g_i(s_i, p_i) := f_i(s_i, p_i) + \max\limits_{s_{i-1}, p_{i-1}} g_{i-1}(s_{i-1}, p_{i-1})$
8 $bp_i(s_i, p_i) := \operatorname*{argmax}\limits_{s_{i-1}, p_{i-1}} g_{i-1}(s_{i-1}, p_{i-1})$
9 **end**
/* Backtracking: */
10 $(s_M^{\text{opt}}, p_M^{\text{opt}}) := \operatorname*{argmax}\limits_{s,p} g_M(s, p)$
11 **for** $i := M$ **downto** 2 **do**
12 $(s_{i-1}^{\text{opt}}, p_{i-1}^{\text{opt}}) := bp_i(s_i^{\text{opt}}, p_i^{\text{opt}})$

Fig. 3: DP algorithm.

restricted by the FSA model (**Fig. 1**(c)) and the constraint (2). The optimized s_i and p_i, denoted as s_i^{opt} and p_i^{opt} respectively, are obtained by the backtracking procedure Step 10–12. Note that the DP algorithm in the proposed technique required about 1.6 sec on a PC (Xeon 1.7GHz) for an image of size 640×100.

2.2 Design of criterion function

The criterion function $f_i(s_i, p_i)$ is designed based on the following observations.

- Long contiguous edges with near-vertical directional feature are often detected around book boundaries.
- Most edges with near-horizontal directional feature are detected around the title characters and the illustrations on book parts.
- Edges are rarely detected in bookshelf background parts and book spine parts.

Thus, the criterion function for the book boundary (i.e., $f_i(q_2, p_i)$ and $f_i(q_3, p_i)$) are designed to take larger value if there are many and/or long near-vertical edges on the line segment at horizontal position i. Since long contiguous edges show the book boundary more clearly than other edges, each edge is given a weight according to the number of the edges connected to it. The criterion function for the book title (i.e., $f_i(q_6, p_i)$) is designed to take larger value if there are many near-horizontal edges on the line segment. The criterion function for the bookshelf background (i.e., $f_i(q_1, p_i)$) is designed to take smaller value if there are many edges on the line segment. The criterion function for the book

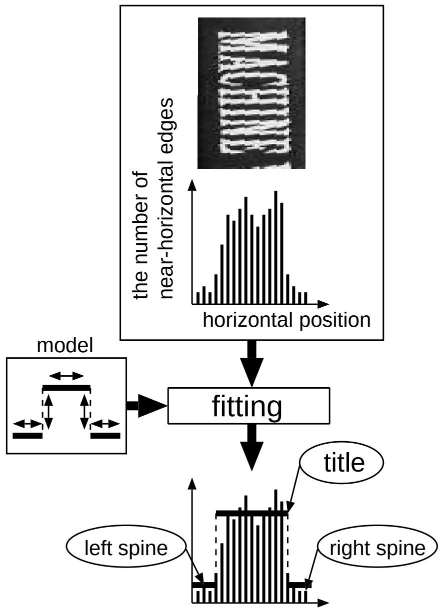

Fig. 4: The post-processing for title extraction. This processing is performed on a book image isolated and slant-corrected.

spine (i.e., $f_i(q_4, p_i)$ and $f_i(q_5, p_i)$) are designed to take smaller value if there are many near-vertical edges on the line segment. For suppressing the effect of the illustrations, near-horizontal edges are disregarded.

3 Title extraction

In this section, the second step of the proposed technique, i.e., title extraction algorithm is presented. The input of this algorithm is the individual book image isolated and slant-corrected using the result of the first step. Note that the title of the book image is already extracted by the first step (recall that our FSA model includes the title state q_6). Unfortunately, the title extraction results by the first step are not sufficient when the boundary detection accuracy is emphasized at the first step. (Namely, it is hard and delicate to detect both boundaries and title parts simultaneously with sufficient accuracy.) Therefore, titles are extracted again at the second step after the accurate boundary detection at the first step.

The following post-processing is performed, which utilizes a projection histogram of near-horizontal edges that are often detected around the title characters. First, for slant-corrected individual book image, the vertical projection histogram of near-horizontal edges is calculated. Second, as shown in **Fig. 4**, the fitting between this histogram profile and a rectangular model is performed. This

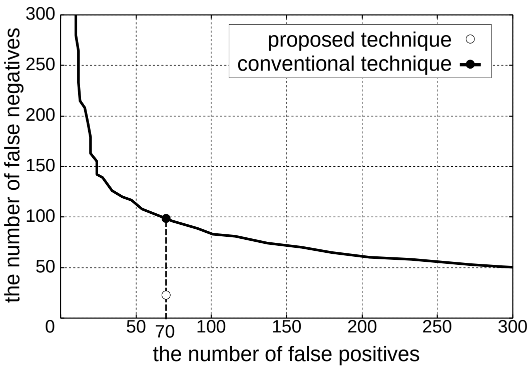

Fig. 5: The numbers of false positives and false negatives. in boundary detection experiment.

fitting is based on the minimization of the mean square error. Finally, the horizontal positions corresponding to the high level part of the model are extracted as a title part.

The above algorithm can be easily extended to cope with titles composed of two or more text lines by assuming additional high level parts. In fact, the extended algorithm was employed in the following experiments.

4 Experimental results

For the experiments to evaluate the proposed technique, 60 bookshelf images were prepared. Then an edge map was obtained by Canny edge detector for each image. All the processes of the proposed technique were performed on this edge map. (Namely, no color feature was used explicitly.)

4.1 Results of boundary detection

The boundary detection performance of the proposed technique was compared with that of the conventional technique [5] using Hough transform. **Figure 5** shows the number of missed boundaries (false negatives) and the number of detected spurious boundaries (false positives). In the graph, "$\circ$" shows the result of the proposed technique ($W = 50$). The evaluation of the conventional technique was conducted while changing the threshold to find local maxima in Hough parameter space, and therefore its result is shown as a curve line in the graph. Hereafter, the result shown by "$\bullet$" was considered as the result of the conventional technique. The number of false negatives of the proposed technique was 3.2% (= 23/727) of all book boundaries and far less than 13.5% (= 98/727) of the conventional technique.

Figure 6 shows several detection results. Book boundaries detected are shown as black lines. It is shown that the proposed technique can correctly detect the boundaries of not only closely stood books but also non-uniformly slanted books. Especially in (b)–(e), it is shown that the proposed technique could detect all boundaries of slanted books while the conventional technique failed.

The proposed technique can not only detect book boundaries but also isolate individual book images from a bookshelf image, whereas the conventional technique can only detect book boundaries. Among the 610 books contained in the 60 bookshelf images, 549 books (90.0%) could be isolated correctly by the proposed technique. Among the 61 books mis-isolated, 31 books were caused by false negatives. The remaining 30 books were caused by false positives.

4.2 Results of title extraction

Figure 6 also shows several title extraction results by the second step of the proposed technique. Title parts extracted are non-shaded. It is shown that the title parts of most books can be extracted appropriately. In (c) and (d), it is shown that the title parts composed of two text lines can be extracted respectively. Through our rough observation, it was shown that most of the title parts (> 90%) could be extracted correctly.

5 Conclusion

A novel technique to extract individual book titles from bookshelf images was proposed in this paper. The proposed technique is composed of two steps. In the first step, the component, such as boundary and spine, and the local slant angle are optimally estimated at each horizontal position of a bookshelf image, and then the individual book images are isolated. For accurate estimation, a finite state automaton model is employed, each of whose states corresponds to a component of bookshelf images. In the second step, the book title is extracted from each book image isolated and slant-corrected using the result of the first step. A vertical projection histogram of its near-horizontal edges is utilized for this purpose.

Experimental results showed that the proposed technique has superior boundary detection performance than the conventional technique. In addition, it was also shown that the proposed technique can appropriately extract the titles of individual books from bookshelf images.

Acknowledgment This work was supported in part by the Ministry of Education, Culture, Sports,

Fig. 6: Examples of book boundary detection and title extraction. From left to right: the original images, the results of the conventional boundary detection technique, the results of the first step of the proposed technique, and the results of the second step of the proposed technique. Book boundaries detected are shown as black lines. A spurious boundary is shown by a red arrow. Boundaries correctly detected only by the proposed technique are shown by blue arrows. In the results of the second step of the proposed technique, title parts extracted are non-shaded.

Science and Technology in Japan under a Grant-in-Aid for Scientific Research No.14780293 and Public Management, Home Affairs, Posts and Telecommunications in Japan under Strategic Information and Communications R&D Promotion Programme (SCOPE).

References

[1] Z. Otake, T. Nagao, T. Agui and M. Nakajima: Extraction of character regions from bookshelf images, Bulletin of the Japanese Society of Printing Science and Technology, vol.29, no.1, pp.38–45, 1992. (in Japanese)

[2] Y. Ishikawa, M. Ohkura, M. Shiono and R. Hashimoto: A study of extraction of character regions from bookshelf images, Bulletin of the Okayama University of Science, vol.32A, pp.163–171, 1997. (in Japanese)

[3] Y. Akiyama and M. Ito: Book recognition from color images of book shelves, Proc. of IAPR Workshop on MVA'98, pp.106–110, 1998.

[4] M. Sawaki, H. Murase and N. Hagita: Character recognition in bookshelf images using context based image templates, Proc. of ICDAR, pp.79–82, 1999.

[5] K. Hiroi, M. Fujiyama, T. Miura and M. Ito: Book boundary extraction from book shelf images, Proc. of the Meeting of Image Recognition and Understanding (MIRU2000), vol.I, pp.271–276, 2000. (in Japanese)

Eiji, Taira: received B.E. and M.E. degrees from Kyushu University in 2000 and 2002, respectively. He is currently a doctoral course student in Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests include pattern recognition. He is a member of IEICE.

Seiichi, Uchida: received B.E., M.E., and Dr.Eng. degrees from Kyushu University in 1990, 1992 and 1999, respectively. From 1992 to 1996, he joined SECOM Co., Ltd., Tokyo, Japan where he worked on speech processing. Now, he is an associate professor at Faculty of Information Science and Electrical Engineering, Kyushu University. His research interests include pattern analysis and image processing. Dr. Uchida is a member of IEEE, IEICE, IPSJ, ITE, and ASJ.

Hiroaki, Sakoe: received the B.E. degree from Kyushu Institute of Technology in 1966, and the M.E. and D.E. degrees from Kyushu University in 1968 and 1987, respectively. In 1968, he joined NEC Corporation and engaged in speech recognition research. In 1989, he left NEC Corporation to become a Professor of Kyushu University. His research interests include speech recognition and pictorial pattern analysis. He received 1979 IEEE ASSP Senior Award, 1980 IEICE Achievement Award, and 1983 IECE Paper Award. He also received Kamura Memorial Prize from Kyushu Institute of Technology. Dr. Sakoe is a member of IEEE, IPSJ, ITE, and ASJ, and a fellow of IEICE.