Paper

Extraction of Text Region From Color Image Sequence of A Book Cover Containing Gilt Characters

Non-member Hua YANG (Keio University)
Member Shinji OZAWA (Keio University)

We have proposed a system for extracting bibliography information from still color image of a book cover. Although it works efficiently for ordinary books, it seems impossible to obtain satisfying results with the exact same system for the book covers containing gilt characters. The reason is that while photographing a book cover onto the image plane, reflection effects on the book cover which contains gilt characters are much more complicated than that on ordinary ones, so that it is impossible to obtain a still image satisfied enough to be processed by the system. When photographing a book cover into a still image, for ordinary book covers there usually only is total diffuse reflection effects. While for those contain gilt characters, there is mixed reflection effects–diffuse reflection and mirror reflection. That mixed effects make some part of book cover invisible in the still image. Instead of a still image, image sequence photographed by a moving camera will be helpful. The reflection on one spatial point is changing because the effects of mirror reflection in different frames. That means that the intensity of the pixel corresponding to that spatial point could be bright enough to be visible in some frames and could be too dark to be visible in others. This paper describes an efficient method for extracting characters from color image sequence of book covers containing gilt characters, taking advantage of reflection changes in different frames to extract enough reflection information for proper character extraction. Integrated with this new method, the proposed system is expected to work more accurately.

Keywords: gilt characters, achromatic region, tracking, correlation, digital library

1. INTRODUCTION

Recent years, digital libraries have been developed increasingly so that it became possible to search bibliographies of books, indexes of journals, and abstracts of papers online. Nevertheless, most of original information is still extracted and input manually, and it requires extensive human labor to build efficient databases. On the other hand, rapid increase in the range and volume of publications makes manual work more and more difficult. Thereby automatic information extraction and input techniques are demanded [1][2], and increasing emphasis is being placed on the realization of computer based systems, which are able to analysis printed documents automatically. A large number of methods for document analysis and automatic information extraction have been proposed. For instance, local linearity algorithm for text-line extraction [3], text string separation from mixed text/graphics images [4], and so on. However, in document analysis researches, much of the attention is paid to the text part [1][2][4]; while the researches in relation to bibliography information extraction are carried out less frequently.

Centered round the idea of automatic extraction of bibliography information from the color image of a book cover, we have proposed a system for extracting text regions and classifying them into proper bibliography categories [5]. However, there left an almost unsolvable problem: gilt characters cannot be extracted and classified correctly. As for the reason of this problem, it is mainly because to each part of the book incident angle is different and such difference makes some parts too bright to be visible or too dark to be visible in one still image. Although there are several method to prevent too strong reflection during the period of photographing, we prefer to adopt the most natural and easiest method, namely to take a sequence of images with a moving camera. This method does not need to prepare too much complex illumination, neither does it hurt any part of the book. The most important and interest thing is that helpful to take advantage of reflection changes in different frames. Meanwhile, taking image sequence with the moving camera makes the resolution improved by taking a part of the book in each frame when the size of the book is wider than field of vision. Unfortunately, there are very few references about gilt character processing.

In this paper, we propose a new method for extracting gilt characters from the color image sequence of a book cover containing gilt characters. This paper is organized as follows. Section 2 describes the framework of this method and its applicability to the system we have proposed; section 3 addresses the problems while extracting text regions from images containing gilt characters; section 4 describes new method to solve the problems addressed in section 3. More experiments and results are shown in section 5 and conclusions in section 6.

2. FRAMEWORK

In the system we have proposed before, a method for extracting bibliography information from still color image of a book cover is described. Although text region extraction is the objective in this system, proper preprocessing is necessary and very important for obtaining satisfying results. Therefore, text region extraction becomes the last part of the system. Because gilt characters make reflection on part of the book cover too complicated to be photographed properly into images, images containing gilt characters were not taken into account at that time.

The objective of this paper is to propose a new method of text region extraction from color images containing gilt characters. By applying this new method, images are processed properly before text region extraction and the results make some parts of the proposed system applicable on images containing gilt characters as well.

2.1 Camera System and Input Images In this system, camera works under the condition shown in Fig. 1, moving through the direction of Y. As a common sense, a book cover consists of three parts: front cover, spine, and back cover. According to this concept, we define each part of the book cover image as shown in Fig. 1. Fig. 2 shows an image taken with such camera system. Color processing is operated on each frame respectively.

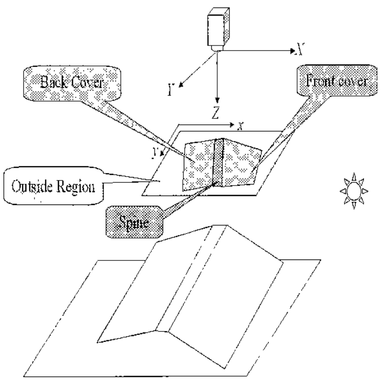


Fig. 1. Camera System

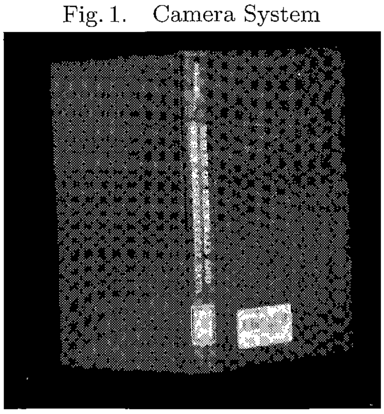


Fig. 2. An Input Image $g(x,y)$

2.2 Process Outline The original input image sequence consists of $\hat{K}$ frames color image, each frame is denoted as $g_k(x,y)$ and $k = 1, 2, ..., \hat{K}$. Fig. 2 shows an example of such image. This system works according to the processing as follows. The details of these processes are discussed in the following sections.

(1) Segmentation of color space. For $g_k(x,y)$, an achromatic image $f_k(x,y)$ and a chromatic image $c_k(x,y)$ is obtained. The histogram functions of $f_k(x,y)$ is denoted by $h_k(f)$.

(2) Requantization of achromatic images. For $f_k(x,y)$, a Λ_k-level requantized image $f'_k(x,y)$ is obtained. It has a key intensity value set $F_k = \{F_k^1, F_k^2, ..., F_k^{\Lambda_k}\}$.

As for the chromatic image sequence, the method we have proposed works effectively.

(3) Tracking and Class Correspondence for requantized images. Q track images are obtained. Considering the convenience for description, we reserve an image $f''_k(x,y)$, which is described by the tracks' number which appear in $f'_k(x,y)$.

(4) Integration of each track image to get Q binary OR images $O^q(x,y)(q = 1, 2, ..., Q)$.

(5) Text region extractions from OR images.

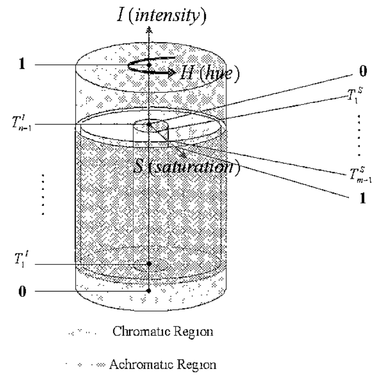


Fig. 3. HSI Color Space Segmentation

2.3 Determination of Achromatic and Chromatic Regions Color space segmentation is one of the first and the most important steps before the extraction and analysis, because any error occurred in segmentation section might influence the extraction of text region [6]. Considering the correspondence between color space and the human visual interpretation of color [7], HSI color space is adopted in this system.

Hue is the most reliable attribute among the three attributes of HSI space, because it is independent of the intensity attribute. However, hue is unstable at very high and very low saturations or very high and very low intensities [6]. To solve this problem, it is necessary to define the effective ranges of hue, i.e., to define achromatic and chromatic areas in HSI color space. In other words, to determine the ranges where intensity or saturation is very high or very low. It can be achieved by thresholding the global intensity where $I = T_1^I, T_2^I, ..., T_{n-1}^I$ and the global saturation where $S = T_1^S, T_2^S, ..., T_{m-1}^S$. We say intensity is very low if its value is lower than T_1^I and very high if its value is higher than T_{n-1}^I. Similarly, we say saturation is very low if its value is lower than T_1^S and very high if its value is higher than T_{m-1}^S (Fig. 3).

As for a color image, it is necessary for every pixel to determine whether it belong to effective hue range or not, namely whether it is chromatic or achromatic. A pixel is defined as achromatic if its intensity is lower than T_1^I or higher than T_{n-1}^I, or its saturation is lower than T_1^S or higher than T_{m-1}^S, the others are defined as chromatic [5]. Achromatic pixels are completely desaturated and their values depend on intensity only; while chromatic pixels have stable hue values and their values depend on all the three attributes.

As a result, an achromatic image (Fig. 4) consists of achromatic pixels with their intensity and a chromatic image consists of chromatic pixels with their value of hue are created. The description in this paper pays attention to the achromatic image, in which gilt characters might be contained.

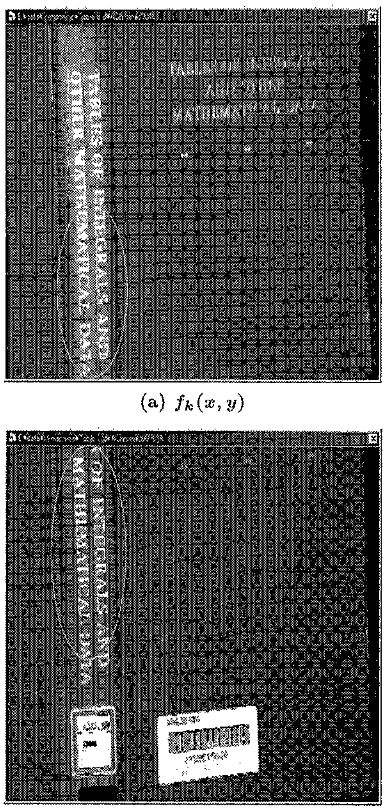
2.4 Achromatic Region Processing Gilt characters come under achromatic region because they are too bright or too dark caused by the reaction to mirror reflection. Therefore, in achromatic region, there are not only achromatic characters in common meaning (effected by total diffuse reflection only) but also gilt characters (effected by both diffuse reflection and mirror reflection). Fig. 4 shows two sample achromatic images, the kth frame and the $k+K$th frame.

Moreover, in this new method, in order to take advantage of reflection angle's changes, a moving camera system is adopted. And to obtain a proper complete image for further processing, special processing of requantization and class corresponding are necessary. Furthermore, when the size of a book cover is wider than visual field, the book cover will be photographed into several frames and only one part of the book cover is taken in each frame– partial photograph. For such cases, the process of frame integration is necessary too. From Section 3 these processes are described in detail.

2.5 Chromatic Region Processing As for chromatic image, it could be processed the same as the system we have proposed, and here will be omitted. However, for a large book cover, some of the parameters obtained when making achromatic frames into one completed image are fine here to make chromatic frames into one suitable image for further processing.

2.6 Text Region Extraction For an image containing text regions, there are two important features.

One is called stroke width, which is defined as the non-unit runlength, the number of that runlength appears a peak in runlength histogram [3]. As shown in Fig. 5, one stroke width corresponds to one kind of font size. When more than one stroke width is detected, the number of stroke width stands for the kind of font. In other words, it is certain that for an image containing text regions there must be at least one stroke width can be detected (Fig. 5), while no stroke width can be detected from an image without text region.



(a) $f_k(x, y)$

(b) $f_{k+K}(x, y)$

Fig. 4. Achromatic Images ($f \in [0, 255]$)

The other important feature is labels, which reflect connected components or pixels, obtained by labeling process. In ideal cases, one label corresponds to one character.

Still there are some common rules when printing characters, that is:

(1) The space between characters within one word must be narrower than the space between different words

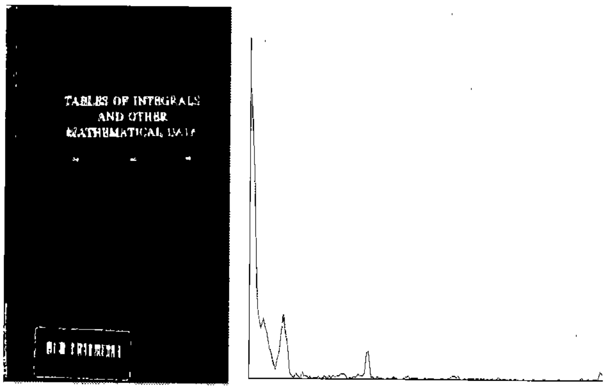
(2) The space between different lines within one paragraph/block must be narrower than the space between different paragraph/block

Applying these rules, an image can be segmented into several blocks. A block can be determined as text block if at least one local stroke width can be detected from it and that/those stroke width(s) is/are (a) subset of the stroke width set detected globally.

With the new method that will be described in detail in following sections as the input, the process of text region extraction we have proposed in the old system [5] can still work well.

3. Problems in Images Containing Gilt Characters

In spite of the conclusion that gilt characters should come under achromatic region, the effects of mirror reflection make some parts of gilt characters invisible in one still image. To solve such problem, continuous images taken by a moving camera are helpful. Because for one part of gilt character, it could be invisible in one



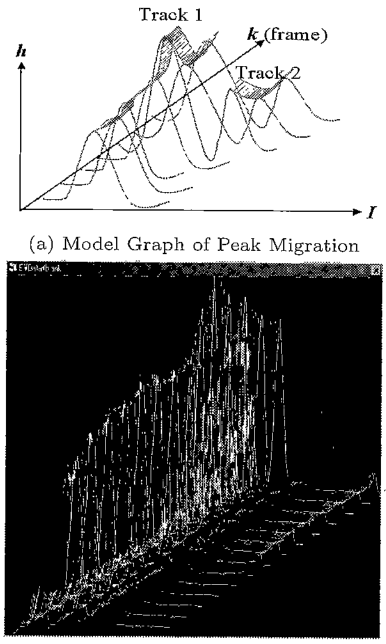
(a) a sample binary image (b) Runlength histogram

Fig. 5. Runlength Histogram

frame but visible in others, for the camera's movements cause lighting direction onto the same spatial point different in different frame. It is assumable that a completed image in which all the gilt characters become visible by joining the pixels that correspond to the same spatial point of the images.

Because the camera keeps moving during photograph period, the coordinate of the same spatial point in different frames will also be different. Therefore it is necessary to track the coordinate changes of each spatial in each frame, i.e. to get the information of camera's movement. Although correlation will be one of solutions, correlation is often operated on binary images. Dynamic binarization seems to be able to make satisfying results, it will be found from the description in the following sections that it is not only a proper binary image desired in this system, it is also desired to find out the correspondence of each class in images to a part of the object in spatial space. For this reason, traditional binary images correlation is not enough here. Requantization based on intensity histogram for each frame is essential. The correlation processing will be operated on one proper class.

Still more, requantized intensity information is inadequate for the tracking – class corresponding is necessary. By this step, labeling process will be put into use in common system. However, for the same reason mentioned above, the area of the same character in different frame will be extremely different. Therefore, tracking the same spatial point within the image sequence is necessary. On the other hand, when photographing range become narrower, some characters will appears suddenly or disappears suddenly so that the tracking in time domain is too difficult to be approached. We propose a method to achieve the tracking in histogram domain, Fig. 6(a) shows the model graph of this method. For a complete-diffusive image sequence, the shape of histogram of each frame, the position of peaks in each histogram and even the altitude of each peak should be steady independent of the camera movement. However when there appears mirror reflection during photograph process, such as during photographing book covers with gilt characters, the number of pixels vary with the movement of camera, so that the position of peaks in each histogram and the altitude of each peak will migrate with that movement too, while the shape stay steady. Nevertheless this migrate range between two frames is narrower than the distance between neighborhood peak even within the same histogram. Accordingly, even there would be some parts appear or disappear suddenly, the peak of histograms will never disappear. This condition makes the tracking in histogram domain possible and it will be simpler and more robust than tracking in time domain. In the model graph there are two tracks obtained as a result of the tracking. Fig. 6(b) shows a real instance.



(a) Model Graph of Peak Migration

(b) Real Instance of Peak Migration

Fig. 6. Peak Migration

In order to apply the text region extraction process proposed in our old system [5], a completed binary image is necessary. We call such a binary image integrated with the method that will be described in 4.3 and 4.4 *OR* image.

In Section 4, method of requantization for segmenting similar parts into the same level, class corresponding for tracking the same part in spatial space, measurement of camera movement for getting camera moving speed between frames and image integration for obtaining a proper image for text region extraction are described in detail.

4. METHOD

4.1 Requantization in Each Frame Pixels those are determined as achromatic are completely desaturated and their values depend on intensity only so that for each achromatic frame of image obtained from color processing, it is possible to be requantized into Λ levels by thresholding their intensity histogram respectively using the fourth central moment method [8] we have proposed. If there includes intensity which belongs to different categories within a range of intensity, the value of fourth central moment within that range is large. On the contrary, should all intensity within a range belongs to one category, the value of fourth central moment within that range is small.

Assuming that $p(f)$ is the normalized histogram of image $f(x,y)$ and $f(x,y)$ can be requantized into up to L levels, P_r is the ratio of the rth category, av_{lr} is average intensity within the rth category when requantizing $f(x,y)$ into l levels and M_{lr} is the fourth central moment of the rth category, the threshold set of $T_l = \{0 = T_{l0}, T_{l1}, ..., T_{lr}, ..., T_{ll} = 255\}$, which meets Eq. 3, is the proper threshold set when requantizing $f(x,y)$ into l levels.

$$P_r = \sum_{f=T_{l(r-1)}}^{T_{lr}-1} p(f)$$

$$av_{lr} = \frac{1}{P_r} \sum_{f=T_{l(r-1)}}^{T_{lr}-1} f \times p(f) \quad \cdots\cdots\cdots\cdots \quad (1)$$

$$M_{lr} = \frac{1}{P_r} \sum_{f=T_{l(r-1)}}^{T_{lr}-1} (f - av_{lr})^4 \times p(f) \quad \cdots\cdots \quad (2)$$

$$EE_l(T_{l1}, ..., T_{lr}, ..., T_{l(l-1)}) = \min_{T_{lr} \in T_l} \sum_{i=1}^{l} M_{lr} \quad \cdot \quad (3)$$

As a result, for each l ($l = 1, 2, ..., \Lambda, ..., L$) one set of threshold T_l is obtained with the minimum error when assuming that $f(x,y)$ is requantized into l levels, and a set of error value EE is obtained.

$$EE = \{EE_1, ..., EE_\Lambda, ..., EE_L\} \quad \cdots\cdots\cdots\cdots \quad (4)$$

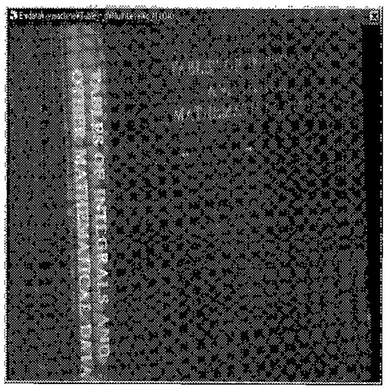
Furthermore, the threshold set of T_Λ is the proper solution for requantization process, where Λ makes $EE_\Lambda = min(EE)$. Thus, both of the reqantization level and threshold set are determined automatically. The image $f(x,y)$ is requantized into Λ levels with the intensity value F^r, which has the maximum pixel number between $T_{\Lambda(r-1)}$ and $T_{\Lambda r}$. Where

$$T_\Lambda = \{T_{\Lambda 0}, ..., T_{\Lambda r}, ..., T_{\Lambda\Lambda}\} \quad \cdots\cdots\cdots\cdots \quad (5)$$

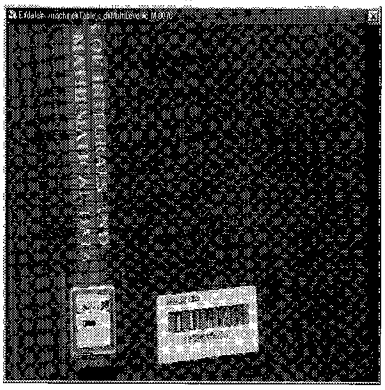
$$F = \{F^1, F^2, ..., F^r, ..., F^\Lambda\} \quad \cdots\cdots\cdots\cdots \quad (6)$$

As shown in Fig. 7, image $f(x,y)$ (Fig. 4) is requantized based on its intensity histogram $h(f)$ (Fig. 8(a)(b)). And with the intensity set of F, a requantized image $f'(x,y)$ (Fig. 7(a)(b))is created; its histogram is $h'(f')$(Fig. 9(a)(b)). Requantization process is operated on each frame so that there is a requantized image sequence created and the following class corresponding process works based on this requantized image sequence.

4.2 Tracking And Class Corresponding Ordinarily tracking each label with labeled image sequence is a common solution for tracking process. However with the image sequence containing gilt characters, visible pixels in one label vary with frames so that the area of one label varies with frames. That makes label correlation cannot estimated properly and consequently makes label tracking very difficult or even hardly possible. In that case it seems possible to track each pixel with a binary image sequence. Unfortunately one spa-

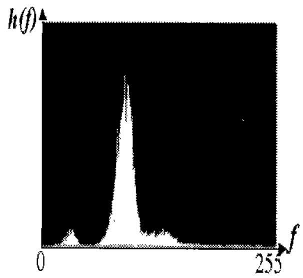


(a) $f'_k(x,y)$

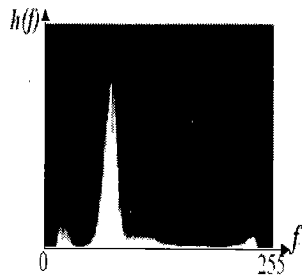


(b) $f'_{k+K}(x,y)$

Fig. 7. Requantized Images ($f' \in F$)

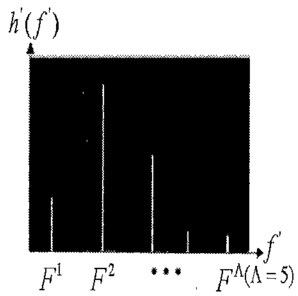


(a) $h_k(f)$

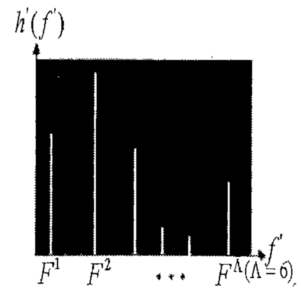


(b) $h_{k+K}(f)$

Fig. 8. Histogram of achromatic images



(a) $h'_k(f')$



(b) $h'_{k+K}(f')$

Fig. 9. Histogram of requantized images

tial point could be visible in some frames and invisible in others so that its track will break halfway. As a result neither is pixel tracking with binary image sequence possible.

In this paper we propose a new method of class tracking in histogram domain with requantized image sequence. As mentioned again and again, the movement of camera makes reflect angel various to different frames and such difference makes one spatial point visible in some frames but invisible in others. In other words, this phenomenon is reflected to histogram with variable area, number and altitude of peaks(Fig. 8(a)(b)). However no matter how area, numbers and altitude of peaks in histogram vary, it is sure that there must appear at least one peak. Hence it will get equivalent effects to track each peak with class tracking.

Moreover, for a partial photographed image sequence there could be some parts appear or disappear in some frames(Fig. 9(a)(b)). In order to get the start frame, end frame and the relationship between new track and existed tracks, class corresponding is a necessary process.

Class j in the $(k+1)$th frame is defined to be correspondence to class i in the kth frame if j could make Eq. 7 the minimum.

$$D_{ij}{}^2 = (F_k^i - F_{k+1}^j)^2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots (7)$$

Thus Q chains of correspondence classes can be obtained within one image sequence and they are called tracks. Each track has a track number. A correspondence track image sequence is created by replacing the class number in requantized image sequence with its correspondent track number (Fig. 10).

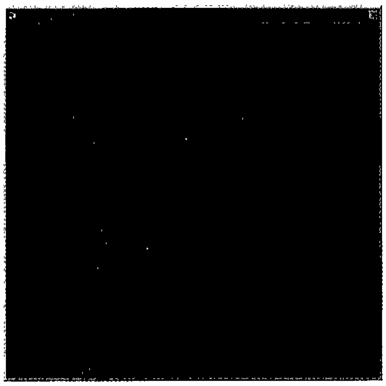
4.3 Measurement of Camera's Movement For each frame of the track image sequence, a pair of (dx, dy) that stands for camera's moving speed between two frames can be estimated by correlation process. $(dx(k), dy(k))$ that makes Eq. 8 the minimum is the result between the kth and $(k+K)$th frame.

$$\Delta^q_{k,k+K} = \sum_{y=0}^{H-1}\sum_{x=0}^{W-1}(1-\delta(f_k''^q(x,y) - f_{k+K}''^q(x',y'))) \quad (8)$$

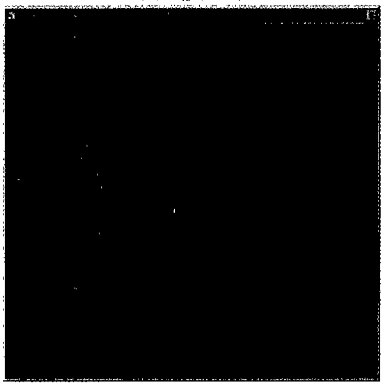
Where
$x' = x + dx$
$y' = y + dy$
H : height of each frame
W : width of each frame
K : frame interval for correlation process

This process is operated on each track respectively to obtain one set of $(dx^q(k), dy^q(k))$ for the qth track. Since the track image sequence is created from one image sequence, there must be one set of $(dx(k), dy(k)$ suitable for all the tracks, in spite of subtle difference between them caused by the effects of requantization.

4.4 Integration Within the Same Track For each track q within the correspondent image sequence, a completed binary image–OR Images can be obtain by the following process (Eq. 9 - 11: OR operation) with the result of $(dx(k), dy(k))$.



(a) $f_k''(x,y)$



(b) $f_{k+K}''(x,y)$

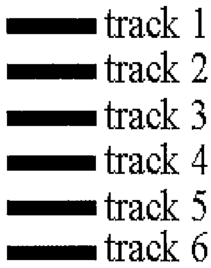


Fig. 10. Track Images ($f'' \in [1, Q=6]$)

$$DX(k) = \sum_{k'=1}^{k} dx(k') \cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

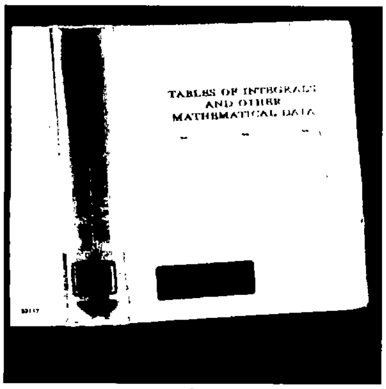
$$DY(k) = \sum_{k'=1}^{k} dy(k') \cdots\cdots\cdots\cdots\cdots\cdots\cdots (10)$$

$$O^q(x,y) = \sum_{k'=K1}^{K2} f_k''^q(x+DX(k), y+DY(k)) \quad (11)$$

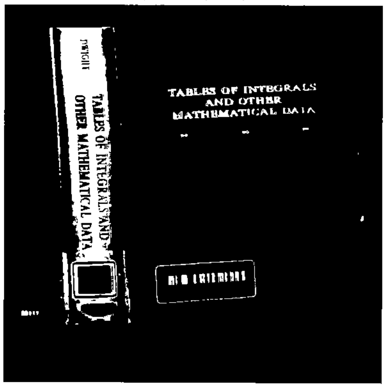
Where $O^q(x,y)$ is the output OR image for the qth track.

However, when there are too many pixels participate the OR operation, the stroke of characters in the result image will be too fat to be processed further. To prevent such situation, suitable range of OR operation $(N1, N2)$ for one pixel need to be determined. There will be Q such OR images created. In Fig. 11 some of the results are shown.

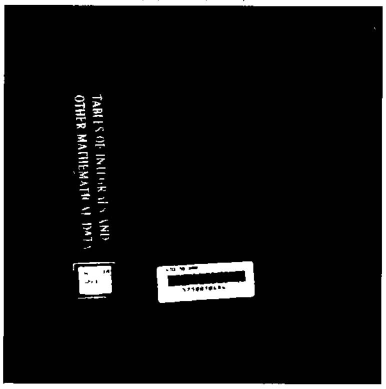
4.5 Geometrical Correction – Affin Transform When an opened book is photographed facedown into an image plane, there must appear some distortion, so that geometrical correction is a necessary processing in such cases. In this system, Affin transform would be enough. Parameters for Affin transform can be obtained with one proper OR image (such as Fig. 11(a)) and the parameters are suitable for other OR images as well. Meanwhile, a book cover consists of front cover, spine, and back cover. As a part of Affin transform effects, front cover and spine, which are our extraction objects, can be segmented at the same time(Fig. 12). They are



(a) $O^2(x, y)$

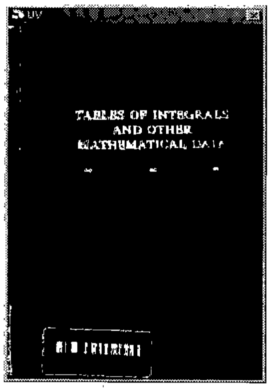


(b) $O^3(x, y)$

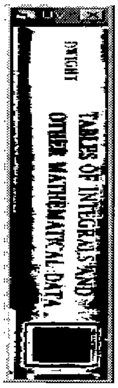


(c) $O^5(x, y)$

Fig. 11. Completed Image of Each Track – *OR* Images

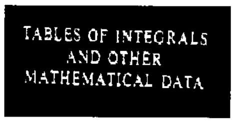


(a) Front Cover



(b) Spine

Fig. 12. Result of Affin Transform



(a) Text Region Extracted from Front Cover

(b) Text Region Extracted from Spine

Fig. 13. Result of Text Region Extraction

Table 1. Experimental Result Data ($\hat{K} = 100$)

No.	Range of Λ	Q	Size of $O(x, y)$	Number of Text Region (T/F)
Fig. 14	3-4	6	656 × 1415	2 (1/1)
Fig. 15	2-3	5	752 × 938	2 (2/0)
1	2-5	6	692 × 729	4 (3/1)
2	3-5	5	651 × 681	2 (2/0)
3	3-5	5	691 × 1409	1 (1/0)
4	2-4	3	669 × 955	3 (3/0)

considered to be equivalent to the binary images for text region extraction with the proposed method [5].

5. EXPERIMENTS AND RESULTS

The image sequences for our experimental study are taken by an equipment called X-Y stage, which controls the camera movement, with a fixed illuminant. The X-Y stage control software is implemented on MS Windows platform. Two hundred frames can be saved during one photographing process and the size of each frame is saved as a 640×480 pixel color image. Camera moves at a speed which can garantee the quality of photographing, and there is a up to 5 pixel difference in y direction and less than 2 pixel difference in x direction between two frames. In our experimental studies, we assume $L = 15$ and $\hat{K} = 100$.

In this section, the experimental results are shown in Fig. 14 and 15. Both the book covers shown in Fig. 14 and Fig. 15 have large size that wider than field of vision. The results of these two experiments show the proposed method is effective for large size objects. The size after track integration and Affin transform of each image sequence is 656 × 1415 pixels and 752 × 938 pixels, respectively. Unfortunately, because of the limited space, we can only use a table to show more experimental results instead of images.

6. CONCLUSION AND DISCUSSION

A new method for extracting gilt characters properly is proposed based on the system of extraction of bibliography information from color image of book cover. This method takes advantage of reflection changes on different kind of book covers during photographing process. From the experimental results, it can be found that this new method is effective for extracting characters from book covers even though there are gilt characters, on which reflection effects are much complicated. And it is effective also for larger size book covers by taking one part of the book cover in each frame.

However, the class criterion function (Eq. 7) in 4.2

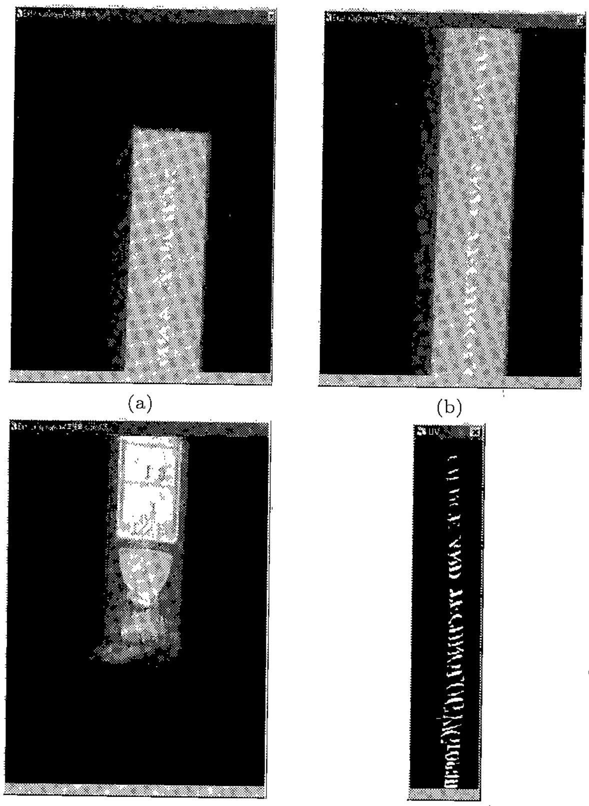


Fig. 14. (a)(b)(c):some frames of image sequence; (d): result of text region extraction

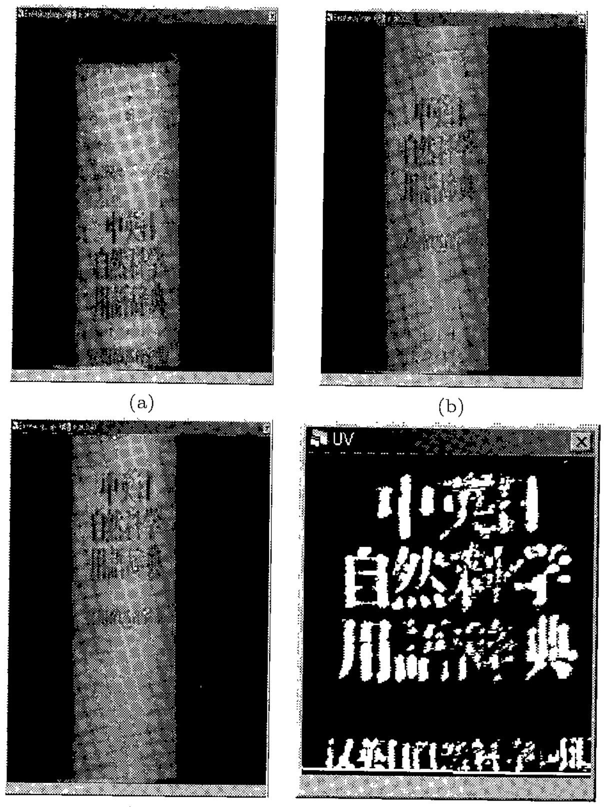


Fig. 15. (a)(b)(c):some frames of image sequence; (d): result of text region extraction

should be defined more accurately as follows. We will work out about how the weight W_p and W_v effect the results and how to determine their value automatically in near future.

$$D_{ij}{}^2 = W_p(h'_k(i) - h'_{k+1}(j))^2 + W_v(F_k^i - F_{k+1}^j)^2 \quad (12)$$

Where
W_p : Weight of pixel number;
W_v : Weight of value of requantized level;

(Manuscript received October 2, 2000, revised January 15, 2001)

References

(1) Kazumi Iwane, Masaki Yamaoka and Osamu Iwaki "A Functional Classification Approach to Layout Analysis of Document Images" Proceedings of the Second International Conference on Document Analysis and Recognition, p774-777, Oct. 20-22, 1993

(2) Shuichi Tsujimoto and Haruo Asada, "Major Components of a Complete Text Reading System", Proceedings of The IEEE, Vol. 80, No. 7, July 1992

(3) Hideaki Goto, Hirotomo Aso "Robust and Fast Text-line Extraction Using Local Linearity of the Text-line" Systems and Computers in Japan, Vol. 26, No. 13, 1995 p327-352, 1989

(4) Lloyd Alan Fletcher and Rangachar Kasturi "A Robust Algorithm For Text String Separation from Mixed Text/Graphics Images" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.10, No.6, Nov., 1988

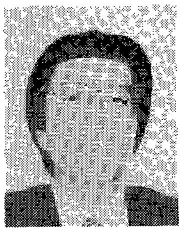
(5) Hua YANG and Shinji OZAWA *"Extraction of Bibliography Information Based on the Image of Book Cover" IEICE Trans. on Information and Systems*, **Vol.E82-D**, No.7, p1109-1116, July 1999

(6) Din-chang Tseng, Chung-hsun Chang "Color Segmentation Using *UCS* Perceptual Attributes" Proc. of National Science Council, Republic of China, Part A, [Physical Science and Engineering] Vol.18 Iss.3 p305-313

(7) I. Pitas, P. Kiniklis "Multichannel Techniques in Color Image Enhancement and Modeling" IEEE Trans. Image Process Vol.5 No.1 p168-171 1996

(8) Sumei Guo, Shinjin Ozawa "A New Criterion Based on Central Moment for Automatic Multilevel Thresholding" IEICI Transactions on Information and Systems, Vol.J80-D-II,No.5,pp.1322-1325,1997-5

Hua YANG (Non-member) receive the BS degree in computer science from Nankai University, China in 1994 and the ME degree in electrical engineering from Keio University, Japan in 1998, respectively. She is currently a PhD student in the Department of Electrical Engineering at Keio University, Japan. Her research interests include computer vision, image processing and pattern recognition.



Shinji OZAWA (Member) received B.E., M.E., and PhD degrees all in Electric Engineering from Keio University, Tokyo, Japan, in 1967, 1969, and 1972, respectively. He had been Research Associate, Assistant Professor, and Associate Professor from 1970 in the Department of Electric Engineering of Keio University. He is now Professor of Department of Information and Computer Science of the same university. At 1984, he had been Visiting Professor of University of Maryland, USA. He has worked on Digital Communication and Signal Processing of Image and Voice. He is a member of IEICE(Japan), IPSJ(Japan), IEEE(USA) and a fellow of IAPR.