Layout Analysis of Tree-Structured Scene Frames in Comic Images

Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi
Department of Information Science
Faculty of Engineering, Utsunomiya University,
Utsunomiya-shi, 321-8585 Japan
taka@athena.is.utsunomiya-u.ac.jp

Abstract

Today, the demand of services for comic contents increases because paper magazines and books are bulky while digital contents can be read anytime and anywhere with cellular phones and PDAs. To convert existing print comic materials into digital format such that they can be read using the cellular phones and the PDAs with small screens, it is necessary to divide each page into scene frames and to determine reading order of the scene frames. The division of comic images into the scene frames can be considered as a type of document layout analysis. We analyzed layout of comic images using density gradient. The method can be applied to comics in which comic balloons or pictures are drawn over scene frames. In this research, a method for detecting the scene frame division in comic images using the density gradient after filling the quadrangle regions in each image with black is proposed. Experimental results show that 80 percent of 672 pages in four print comic booklets are successfully divided into scene frames by the proposed method.

1 Introduction

In this paper, a technique for searching the order to be read and the position of the scene frame in the page is proposed based on structure analysis for the image of each page of print comic materials.

Today, communication networks are in wide spread use, copyright management systems are in practical use, and many services using digital contents have been started in various fields. The demand of the service for comic contents increases because paper magazines and books are bulky while digital contents can be read anytime and anywhere with cellular phones and PDAs. Actually, some services for digital comics have been started. eBooksJapan[1] is the specialty site which sells the digital books, and the number of selling comics has exceeded 10,000. In the portal site YAHOO! JAPAN, an online delivery service for comics is started by the name of a YAHOO! JAPAN Comics[2]. Digital Comic of Rakuten ICHIBA[3] is selling comics of about 6000 volumes. Many of them provide comic images to users page by page. Some of new comics such as "SALARY MAN KINTARO"[4] are published as the exclusive use of online delivery services. They are displayed on personal computers with respect to each scene frame having BGMs and sound effects. Their presentations are advanced to next frames automatically. "comic i" and "comic cmoa"[5], which are the services intended for cellular phones, have been favorably well received at the point of presenting each scene frame reedited with action effects.

To convert existing print comic materials into digital format such that they can be presented using the cellular phones and the PDAs with small screens, it is necessary to divide each page into scene frames and to determine reading order after converting each page into digital form using a scanner.

It is very convenient not to have to use any keyboard or mouse when personal computer users would enjoy comic contents lying sprawled.

As a method for providing such lazy viewing to us, automatic page turning by eye-gaze tracking is considered. In order to implement the automatic page turning, it is necessary to detect the timing when a user finishes the reading of a current page. To detect the timing, it is considered that a eye-gaze tracking system can be used. For that purpose, it is necessary to know the arrangement of scene frames and the reading order in the page.

The division of comic images into the scene frames can be considered as document layout analysis for comic images. In document layout analysis [Ishitani, 2001], the shape of region of interest is assumed to be only rectangle in most cases, while shapes of scene frames in comic images are fundamentally quadrangle which is not limited to rectangle. Therefore, a new analysis method for comic images is needed.

We find that the most of comic pages can be divided into scene frames by recursive guillotine cut, and that the process of the guillotine cut makes the tree structure of the scene

[1] eBOOK Initiative Japan Co.,Ltd.,
http://www.ebookjapan.jp/shop/

[2] Yahoo Japan Corporation, http://comics.yahoo.co.jp/

[3] Rakuten,Inc., http://dl.rakuten.co.jp/shop/rt/comic/

[4] author Hiroshi Motomiya,
http://dl.rakuten.co.jp/shop/rt/sp/kintaro/

[5] NTT SOLMARE CORPORATION,
http://dl.foobio.net/free/cmoa/index.html

frames which has an information to determine the order of the scene frames to be read. The concept of guillotine cut is used by optimization of cutting a sheet into quadrangles requested [Ono, 1999].

As a layout analysis method of comics, Yamada et al. [Yamada *et al.*, 2004] has been realized an algorithm for division and alignment of the scene frame by tracking white region between the scene frame of the comic image. But, only 23 images of comics pages are used by experiments, since main purpose is a realization of the method of image compression and extraction efficiently considered built-in memory capacity and telecommunications cost of cellular phone. And, when the comic balloon or picture are drawn over scene frame, the division of the scene frame can not be well detected in this algorithm.

We analyzed layout of comic images using density gradient [Tanaka *et al.*, 2006]. The method can be applied to comics in which comic balloons or pictures are drawn over scene frames. Experimental results of the method show that 60 percent of 672 pages in four print comic booklets are successfully divided into scene frames.

In this research, we propose an improved detection method of scene frame division for our previous method by preprocessing comic images before detection of division line. And experimental results of scene frame division for sum total of 672 pages of 4 print comic materials are reported.

2 Comics Structure

2.1 Comics Scene Frame Definition

In this study, the scene frame of comics is defined as a quadrilateral frame composed of four straight lines including picture in its interior. Then, the frame lines interrupted by a balloon or a picture are also regarded as scene frame. In addition, the scene frame that consists of imaginary quadrilateral frame line made by other scene frame or the edge of the page is also contained. Though scene frames of comics are mere frame in single panel cartoons, scene frames of comics represent time and space in multiple panel cartoons [Takekuma and Natsume, 1995]. In Japanese comics, reading order is fundamentally from right to left and downward. Reading order is uniquely decided by the arrangement of scene frames.

2.2 Comics Tree Structure

The structure of a comics page can be represented by a tree, whose root is the whole comic image (see Figure 2). The tree is obtained by dividing the image recursively. An example of the structure of comics represented by a tree is shown in Figure 1.

In the example, firstly the comic image is divided into three frames by horizontal division lines in left part of Figure 3 (gray lines are these horizontal division lines), and three leaves are made under the root in right part of Figure 3.

Secondly, the top of the frames (the left leaf which is the elder son on layer 1 in left part of Figure 4) is divided into right and left frames by a vertical division line in left part of Figure 4, and two leaves are made under the node of the elder son in right part of Figure 4.

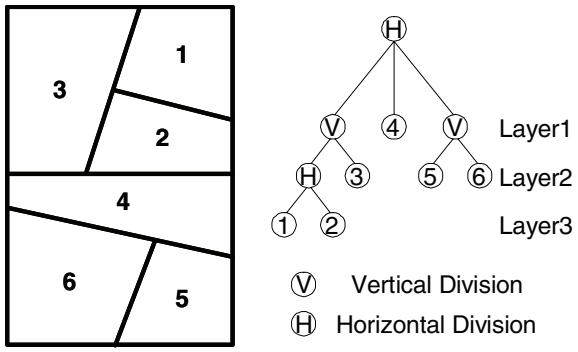

Figure 1: The structure of a comics page (left) and the tree structure of it (right).
Figures within scene frames in comics page shown reading order, and this reading order is shown by the depth first search of this tree.

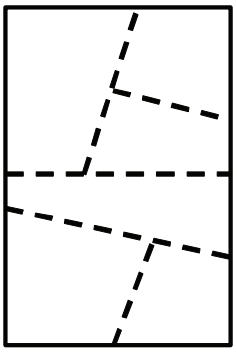

Figure 2: The structure of a comics page before division into scene frames (left) and the tree structure of it (right).

Thirdly, the right frame is divided into top and bottom scene frame, and two leaves are made under the node.

Finally, the bottom of three frames divided into firstly is divided into right and left frames, and two leaves are made at the node.

In this way, the tree in right part of Figure 1 is created. The order of scene frames is decided by ordering the leaves of this tree by means of the depth first search. Thus, it is shown that the order of scene frames of the comics is decided by forming the tree structure.

2.3 Structure of the Scene Frame out of Scope in Our Proposed Method

In the case where scene frames exist inside of a scene frame, the page can not form the tree structure by dividing the image recursively. Figure 5 shows an example of such page.

In such case, it is considered one scene frame including the scene frame which exists in the inside. If such scene frame is the whole image of a page, the scene frame is not chosen as the object of experiments. And if such scene frame appears in the stage on the way of the scene frame division, the scene frame division is not subsequently carried out.

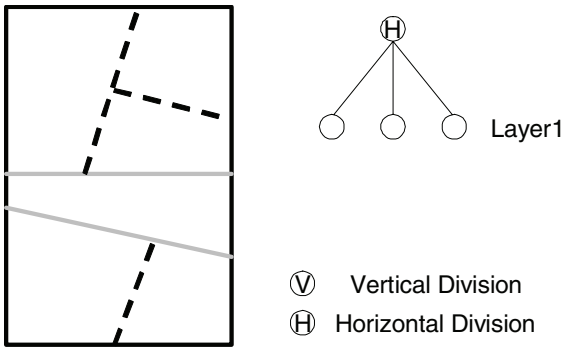

Figure 3: The structure of a comics page at first step (left) and the tree structure of it (right).

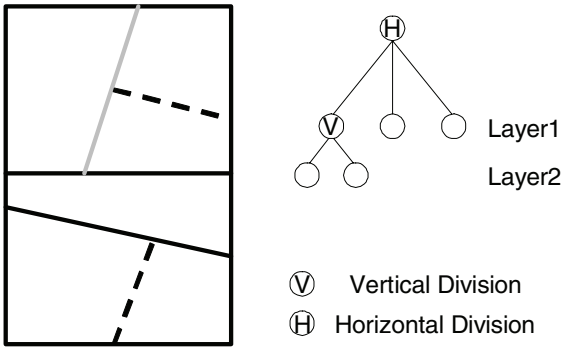

Figure 4: The structure of a comics page at second step (left) and the tree structure of it (right).

3 Division Line Detection Using Density Gradient

3.1 Representation of Straight Line

A straight line is represented by ρ which is the length of the perpendicular line from the origin to the straight line and θ which is the angle between the X-axis and the perpendicular line (see Figure 6). Here, the origin (0, 0) is the center of comics image, the x-axis is defined that the rightward direction is positive and the y-axis is defined that the downward direction is positive in a comics page image.

The straight line $L(\rho, \theta)$ are expressed as follows:

$$\rho = x\cos\theta + y\sin\theta \tag{1}$$

3.2 Weighted Accumulation Value of Gradient

The density gradient $g_\theta(x, y)$ of the θ direction at the position (x, y) on an image is given as follows:

$$g_\theta(x, y) = g_x(x, y)\cos\theta + g_y(x, y)\sin\theta \tag{2}$$

where $g_x(x, y)$ is the density gradient of the X-axis direction, and $g_y(x, y)$ is the density gradient of the Y-axis direction.

The accumulation value $A(\rho, \theta)$ of the density gradient value of the θ direction is obtained along the straight line

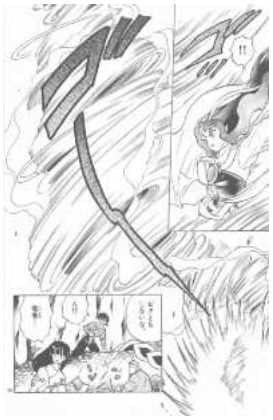

Figure 5: An example page out of scope in the proposed method (source: title Ranma 1/2, author Rumiko Takahashi, publisher Shogakukan Inc., volume 38 p.108).

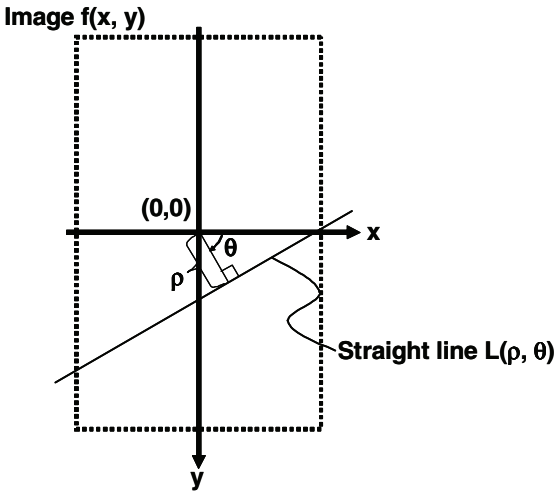

Figure 6: Representation of straight line $L(\rho, \theta)$.

$L(\rho, \theta)$. The set of pixel positions on the straight line $L(\rho, \theta)$ is defined as $S(\rho, \theta)$. The density gradient value $A(\rho, \theta)$ of the straight line $L(\rho, \theta)$ on an image is given as follows:

$$A(\rho, \theta) = \sum_{(x,y)\in S(\rho,\theta)} g_\theta(x, y) \tag{3}$$

The central part of the image is expected that there is the high possibility a division line is located. Therefore, in order to raise the weight of the central part of the image, a weighted accumulation value of gradient $A_W(\rho, \theta)$ is obtained by multiplying $A(\rho, \theta)$ by a Gaussian function $G(\rho)$ on ρ as shown in Figure 7.

$$A_W(\rho, \theta) = G(\rho)A(\rho, \theta) \tag{4}$$

where

$G(\rho) = \exp(-\rho^2/\sigma^2)$, and $\sigma =$ [Image height]$/4$.

We use an evaluation value based on the $A_W(\rho, \theta)$ to search division lines.

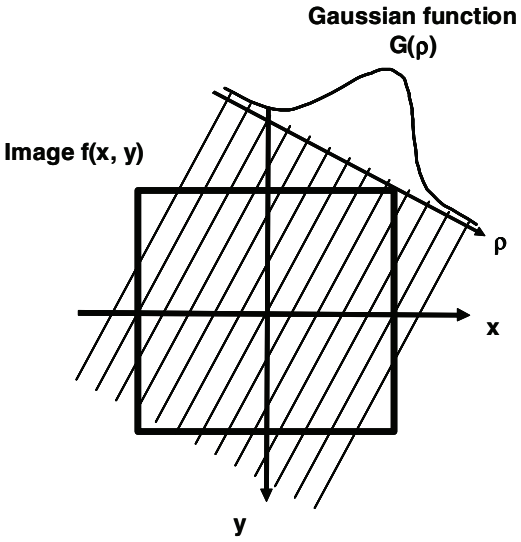

Figure 7: Weighted Accumulation Value of Gradient.

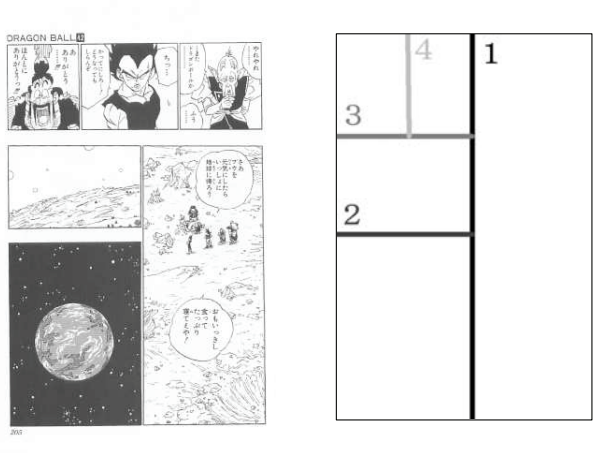

Figure 8: Input image (source: title DRAGON BALL, author Akira Toriyama, publisher SHUEISHA Inc., volume 42 p.205) (left), and the scene frame division image for input image using the convetional method (right).

3.3 Continued White Pixels

Since scene frames are separated by a white pixel region, white pixels continued to some extent length exist in the original image near division lines detected by the density gradient. Therefore, the constant width (the width ± 10 pixels) of the original image is scanned at the same angle and white pixels continued to some extent length are searched after the division line was detected from the density gradient image. Then, the evaluation value $H_W(\rho, \theta)$ of segment without white pixels of some extent length is lowered by multiplying $A_W(\rho, \theta)$ by the ratio of the length of the longest white pixel column to that of the segment. Here, the threshold of the white pixel is set to 224 in 256 tones of gray scale.

3.4 T-type Division Line Judgment

In our previous method, the failure showed the side effect that can be applied to the case where comic balloons or pictures overlap with scene frames. For example, when the shape of the region of interest is vertically extending and its correct division lines are T-type, the division of it by the horizontal line should be followed by the division of the lower half by the vertical line. But the vertical division line actually is firstly detected because of the superiority of the accumulated gradient value of the vertical division line over the horizontal one (the division line of layer 1 in Figure 8)

The figures "1" to "3" in Figure 8 are the layer numbers in the tree structure. In order to correct this problem, a processing for judgment of a T-type division line is added. In vertical division lines and horizontal division lines, each division line that the evaluation value $H_W(\rho_p, \theta_p)$ is the highest is chosen. Then, T-type judgment is processed when the ratio between each evaluation value $H_W(\rho_p, \theta_p)$ is less than 3 times.

First, the intersection point of a candidate of the vertical division line and that of horizontal one is detected. Next, in the neighborhood at the intersection point $(40 \times 40 pixels)$ the evaluation values T_v and T_h of the vertical and horizontal lines respectively are calculated. Last, if $T_v > T_h$, $H_W(\rho_p, \theta_p)$ for the vertical division line is increased by multiplying it by the ratio of T_v to T_h, and vice versa, so that the T-type is well divided by the following processing.

3.5 Division Line Decision

The straight line judged as division line is stored in a taboo list. The straight line which is similar to the straight line stored in the taboo list is not chosen as a division line in the matter of subsequently division line detection. Under the constraint, the candidate of division line is obtained as

$$(\rho_p, \theta_p) = \arg\max_{\rho,\theta} H_W(\rho, \theta).$$

Whether $L(\rho_p, \theta_p)$ is division line or not is decided by whether $H_W(\rho_p, \theta_p) \geq h_{th}$ or not. As h_{th} is raised, division lines detected is limited more. We decided h_{th} heuristically. When the range of θ of detected division lines is $-45^\circ < \theta_P \leq 45^\circ$, the division line $L(\rho_p, \theta_p)$ is made to be a vertical division line. When $-90^\circ < \theta_P \leq -45^\circ$ or $45^\circ < \theta_P \leq 90^\circ$, the division line $L(\rho_p, \theta_p)$ is made to be a horizontal division line. In each divided region, the division is repeated by the same procedure until a division line is not detected.

3.6 Implementation

θ is changed from -90° to 90° at 1° step, and ρ is changed within the image at 1 pixel step. We used straight line generation algorithm DDA in the digital image when accumulating the density gradient of each pixel along straight line.

Since the outside frame of a page image should not be detected as a division line, we trim input page images such that they do not include the outside frame in advance.

3.7 Preprocessing

As a preprocessing of the division line detection, we approximate the outlines of input images with polygons. In the

polygonal approximation, we use the Douglas-Peucker approximation [Hershberger and Snoeyink, 1992]. From the approximated polygons we choose quadrangles. Then we fill the region of the quadrangles with black in order to reduce the influence of pictures within scene frames on the evaluation value for division line detection and to avoid detecting division lines within scene frames.

4 Experiments

Experiments were conducted for the images of pages in print comics excluding the pages composed of single scene frames such as title pages, contents pages, and particular pages in main volume. We used here 672 pages in four print comics. Each image of page was scanned in the resolution of 650 by 1000 pixels.

4.1 Experimental Results

Experimental results for the image set of 672 pages in four print comics are shown in Table 1.

Method A in Table 1 denotes our previous method, method B indicated the proposed method without the preprocessing, and method C is the proposed method with the preprocessing.

The results were classified into five groups such as "succeeded", "failed", "-1", "+1", and "the others". The term "succeeded" means the success for the scene frame division. The term "failed" means the failure for the scene frame division. The term "-1" means that one division line is lacking for the correct scene frame division. The term "+1" means that there is one division line over the correct ones. The term "the others" means that there is an excess or deficiency of two division lines and over.

For the left part of Figure 8 as an input image, a result by the proposed method is shown in Figure 9. In this case the proposed method judging T-type division line is succeeded while our previous method is failed.

Table 1: Experimental results for the comic image set of 672 pages. Values separated by " / " within each cell show the number of pages and ratio(%) respectively.

classification of results	method		
	A		C
succeeded	426 63.4	513 76.3	534 7 .5
failed	13.1	47 7.	37 5.5
-1	7 1 .4	4 7.1	46 6.
1	36 5.4	23 3.4	14 2.1
the others	52 7.7	41 6.1	41 6.1
total	672 1	672 1	672 1

4.2 Discussion

In these experimental results, proposed method have a 16% higher success rate than our previous method. By filling the region of the quadrangles with black, detecting division lines within scene frames was reduced. But, when scene frames that comic balloons or pictures are drawn over scene frames and the outline of comic image is part of scene frame, the scene frames are hardly detected as quadrangles. An example of the latter is shown in Figure 10. In the bottom scene frames, two lines have also been detected as a division line.

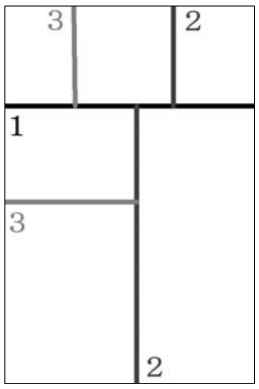

Figure 9: The scene frame division image for left part of Figure 8 using the proposed method.

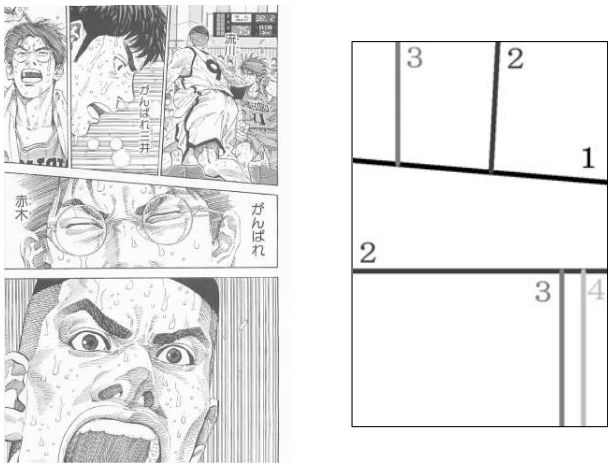

Figure 10: Input image (source: title SLAM DUNK, author Takehiko Inoue, publisher SHUEISHA Inc., volume 31 p.71) (left), and the scene frame division image for input image (right).

Several T-type division lines didn't be detected without the preprocessing. But, by preprocessing, some of them are detected. These examples are shown in Figure 11. The scene frame division succeeded by judging T-type division line using the proposed method.

Since comic drawing style, such as the arrangement of the scene frame, the picture which overlapped with the frame line of the scene frame, and so on, is different by comic artists, the success rate of the experiments is greatly different by each comic. The experimental results for the comic image set of DRAGON BALL volume 42 are shown in Table 2. In the comic image set of DRAGON BALL volume 42, experimen-

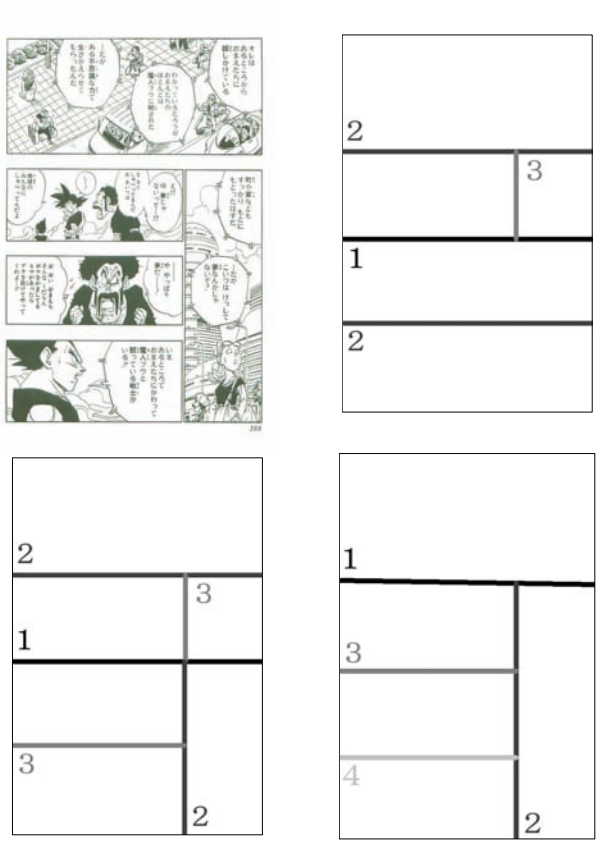

Figure 11: Input image (source: title DRAGON BALL, author Akira Toriyama, publisher SHUEISHA Inc., volume 42 p.205) (upper left), and the scene frame division image for input image using the convetional method (upper right), and the scene frame division image for input image using the proposed method without preprocess (lower left), and the scene frame division image for input image using the proposed method (lower right).

tal results without preprocess have a 32% higher success rate than experimental results of our previous method.

In addition, experimental results of the proposed method have a 8% higher success rate than experimental results without preprocess. The superiority of success rate was remarkable as compared with other three comics.

5 Conclusion

In this paper, a method for detecting the scene frame division in comic images using the density gradient after filling the quadrangle regions in each image with black was proposed and experimental results of scene frame division for the sum total of 672 pages were reported. In these experimental results, the success rate was 80% as a whole, and these experimental results have a 16% higher success rate than experimental results of our previous method. It is thought that it is satisfactory success rate for scene frames including complicated arrangement. Future works are listed as follows:

Table 2: Experimental results for the comic image set of DRAGON BALL volume 42.

classification of results	method		
	A		C
succeeded	111 5 .7	1 2 3.1	1 5 .
failed	3 17.	12 5.5	4 1.
-1	3 13.7	11 5.	3.7
1	13 5.	5 2.3	2 .
the others	26 11.	4.1	1 4.6
total	21 1	21 1	21 1

- We make it possible to correctly divide the failed pages in the experiments here.
- We make it possible to detect the pages with only one scene frames which need not be divided any more.
- In the scene of flashback, the region between neighboring scene frames may be colored by black. We make it possible to correctly divide such regions that the proposed method can not treat at present.

References

[Hershberger and Snoeyink, 1992] John Hershberger and Jack Snoeyink. Speeding up the douglas-peucker line-simplification algorithm. In *Proceedings of the 5th International Symposium on Spatial Data Handling. IGU Commission on GIS*, pages 134–143, 1992.

[Ishitani, 2001] Yasuto Ishitani. Document layout analysis by interaction between data-driven processing and concept-driven processing. *Journal of Information Processing Society of Japan*, 42(11):2711–2723, November 2001.

[Ono, 1999] Toshihiko Ono. Optimizing two-dimensional guillotine cut by genetic algorithms. In *Proc. of the Ninth AJOU-FIT-NUST Joint Seminar*, pages 40–47, July 1999.

[Takekuma and Natsume, 1995] Kentaro Takekuma and Fusanosuke Natsume. *Reading of comics: As it is understood, it can not be explained! Why are comics interesting? Separate volume takarajimaEX (in Japanese)*, pages 168–215. TAKARAJIMASHA,Inc, May 1995.

[Tanaka *et al.*, 2006] Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi. Layout analysis of scene frames in comic images. In *Proceedings of The 13th International Display Workshops (IDW'06)*, VHFp - 20, December 6-8 2006. (accepted).

[Yamada *et al.*, 2004] Masashi Yamada, Rahmat Budiarto, Mamoru Endoo, and Shinya Miyazaki. Comic image decomposition for reading comics on cellular phones. *IEICE transactions on information and systems*, E87-D(6):1370–1376, June 2004.