Frame Segmentation Used MLP-Based X-Y Recursive for Mobile Cartoon Content

Eunjung Han[1], Kirak Kim[1], HwangKyu Yang[2], and Keechul Jung[1]

[1] HCI Lab., School of Media, College of Information Technology,
Soongsil University, 156-743, Seoul, S. Korea
{hanej,raks,kcjung}@ssu.ac.kr
[2] Department of Multimedia Engineering, Dongseo University, 617-716, Busan, S. Korea
hayang88@hanmail.net

Abstract. With rapid growth of the mobile industry, the limitation of small screen mobile is attracting a lot of researchers attention for transforming on/off-line contents into mobile contents. Frame segmentation for limited mobile browsers is the key point of off-line contents tranformation. The X-Y recursive cut algorithm has been widely used for frame segmentation in document analysis. However, this algorithm has drawbacks for cartoon images which have various image types and image with noises, especially the online cartoon contents obtain during scanning. In this paper, we propose a method to segment on/off-line cartoon contents into fitted frames for the mobile screen. This makes the x-y recursive cut algorithm difficult to find the exact cutting point. Therefore we use a method by combining two concepts: an X-Y recursive cut algorithm to extract candidate segmenting positions which shows a good performance on noises free contents, and Multi-Layer Perceptrons (MLP) concept use on candidate for verification. These methods can increase the accuracy of the frame segmentation and feasible to apply on various off-line cartoon images with frames.

Keywords: MLP, X-Y recursive, frame segmentation, mobile cartoon contents.

1 Introduction

As mobile devices are widespread, the request for the use of mobile contents has recently increased and attracts attention in mobile and ubiquitous computing research. For the purpose of satisfying the request, not only producing various contents but transforming on/off-line contents into fitted size is being issue for the mobile devices with a small screen. The transformation of off-line contents has two general types: first is a passive method that person directly transforms the contents, and second is using auto transformation system from off-line to mobile contents. The automatic system can provide the most efficient and convenient method for manipulating and processing the contents. However, these automatic systems have some problems to apply on various contents. Especially to solve some problems of the auto transformation from off-line cartoon contents to mobile contents. The paper-based image of off-line cartoon should

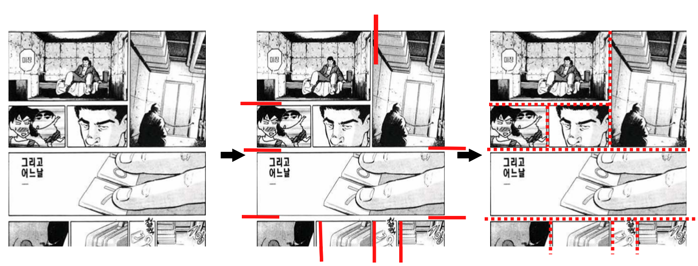
J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 872–881, 2007.

transform to limited size on various mobile browsers. In this case, the key point of transformation the cartoon contents is how to segment the frame of the image effectively in order to fit in small mobile screen.

A number of approaches to page segmentation or page decomposition have been proposed in the literature. Wang et al. [1] used such an approach to segment newspaper images into component regions and Li and Gray [2] used wavelet coefficient distributions to do top-down classification of complex document images. Etemad et al. [3] used fuzzy decision rules for bottom-up clustering of pixels using a neural network. An alternative approach is to use the white spaces available in document images to find the boundaries of text or image regions as proposed by Pavlidis [4]. Many approaches to page segmentation concentrate on processing background pixels or using the “white space” [5-8] in a page to identify homogeneous regions. These techniques include X-Y tree [9-10], pixel-based projection profile [11] and connected component-based projection profile [12], white space tracing and white space thinning [13]. They can be regarded as top-down approaches [14-16] which segment a page recursively by X-cut and Y-cut from large components, starting with the whole page to small components eventually reaching individual characters.

In our previous work [17], we implement frame segmentation[1] for cartoon images, which are one of the most popular ones, with frames, and used X-Y recursive cut algorithm to separate the cartoon contents with frames into frames to fix into small screen mobile device. However, the X-Y recursive cut algorithm has some problems. If the frame boundary has some noises, the algorithm can not segment the frame. The noise data of the image affects the value of the image by projection profile process therefore the method can not detect the frame. And if the frame line is not a straight line, the X-Y recursive algorithm can not make correct segmentation. It can not recognize the line as frame boundary. Due to this reason, the X-Y recursive method can apply to on a limited normal image.

In this paper, we propose an improved method to segment the off-line cartoon frame using the MLP-based X-Y recursive algorithm. The input of the neural network is a scanned image of the cartoon and the output is candidate cutting points of the input image. In this method, several candidate cutting points are first generated by



(a) (b) (c)

Fig. 1. Outline of the proposed method: (a) an input image [19], (b) a result (the gray line denotes boundary) of the forward process, (c) a segmented result

[1] It is a prerequisite stage for extracting important information (salient regions).

X-Y recursive cut algorithm, and we can identify whether the point indicates the right position using the MLP-based segmentation process on only candidate cutting points per each step (Fig.1).

2 Frame Segmentation

We use the input data from the scanned image of off-line cartoon, and change it the binary image (Fig.2). In the binary image of the input data, the black and white pixels are recognized values in the MLP-based frame segmentation process. The input data makes the cutting point by MLP-based segmentation process. In this process, the MLP use the weight which is calculated by training process of some input images. The MLP finds the position of frame segmentation for the cartoon image. Then we can use the cutting-point for the segmenting position and segment the frame using X-Y recursive concept. If the results have two or more candidate points, we could choose the right point by the verification process using projection profile method. In chapter 3, it will be explained the MLP structure and the cutting-point marking process.

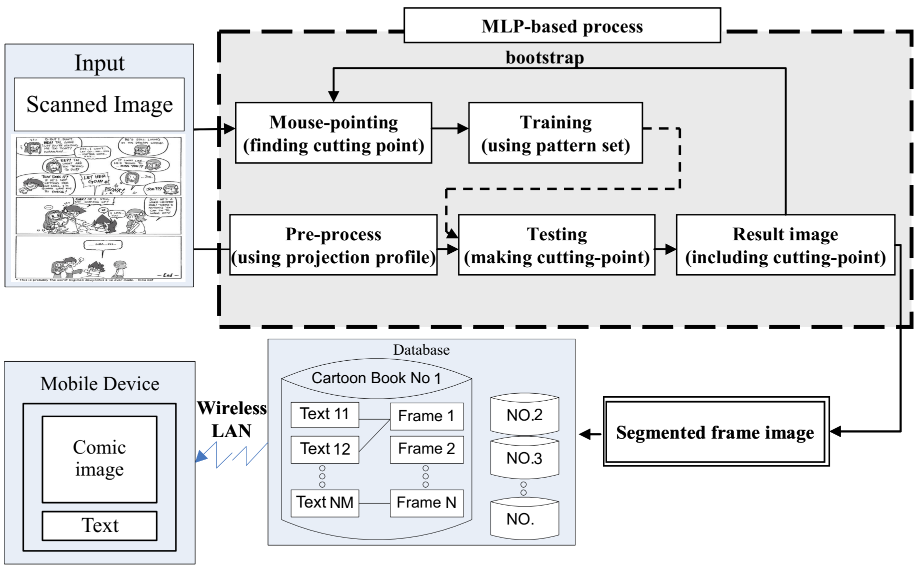


Fig. 2. Overview of the Proposed Approach

2.1 Pre-process

We use the projection profile method for input image of the cartoon to make the input of the Testing process. As shown in Fig. 3, the histogram of the image is used to find the candidate cutting-point area using loose threshold value. Then, the position from the x-y projection profile indicates the input to the Testing process. It has less input value than whole input from the image that can be more efficient processing time.

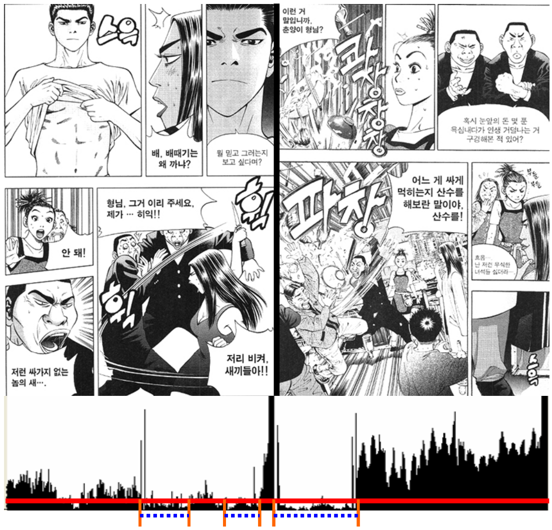


Fig. 3. The input area of Testing process (dotted line indicates the input area)

2.2 Structure of MLP

The structure of the MLP of our proposed design consists of 48 input nodes, 40 hidden nodes, and 1 output node. Fig. 4 shows the structure of two-layer neural network.

It has a fully-connected structure and uses a back-propagation learning algorithm. The MLP inputs a 48-order mesh vector to the network, which is extracted from a 30×40 normalized binary image. The 48 integer values are obtained by counting the number of pixels in each 5×5 local window in the normalized binary image. The resulting 25 intensities are normalized to the range of [0.0 .. 1.0]. Forty-eight floating point numbers are then input into the network in column major order. If the position of the cartoon image is clicked by the mouse, the MLP recognize it frame boundary to

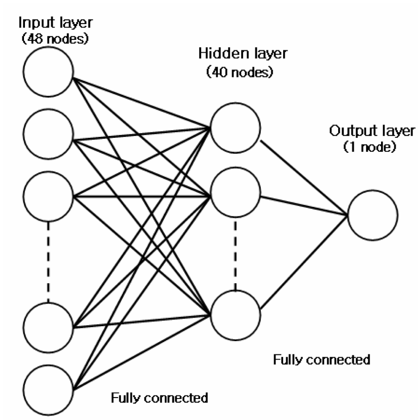


Fig. 4. Structure of two-layer neural network

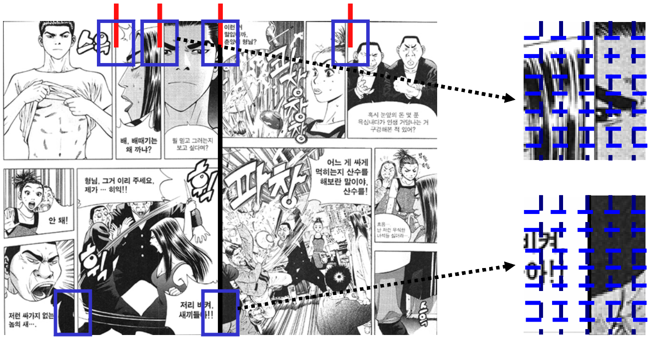


Fig. 5. The MLP input data of the cartoon image and zoomed in images

segment the image. A desired cutting point is determined manually and saved behind the 48-order mesh vector. Fig. 5 shows the process of obtaining neural input values.

The input values from the image are obtained to the boundary area of the images as a quadrangle from left to right and from top to bottom. The output value of the MLP is 1 or 0. In the forward process, the input image which is analyzed into 30×40 input pixels that made by 48-nodes would have the MLP results. The true value indicates the frame boundary. If the area of the 30×40 input has true value, the process can marking the cutting point. The position of the cutting-point takes a role of segmenting point. If the result is false the process can recognize that the area of the 30×40 input is not the frame boundary.

2.3 Cutting-Point Marking

The MLP can find the frame boundary and segment the frame. Fig. 6 shows the cutting-point area.

The line of the frame boundary indicates the cutting-point. The process can recognize the segmentation area with the MLP results. As you can see in the Fig. 6, it is

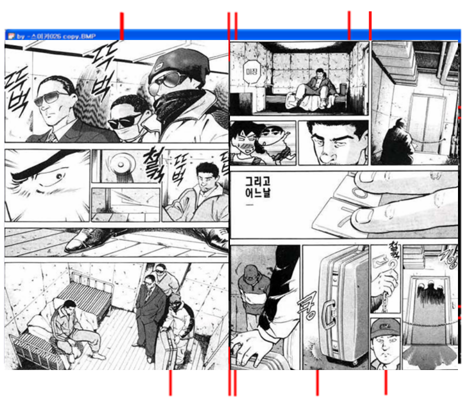
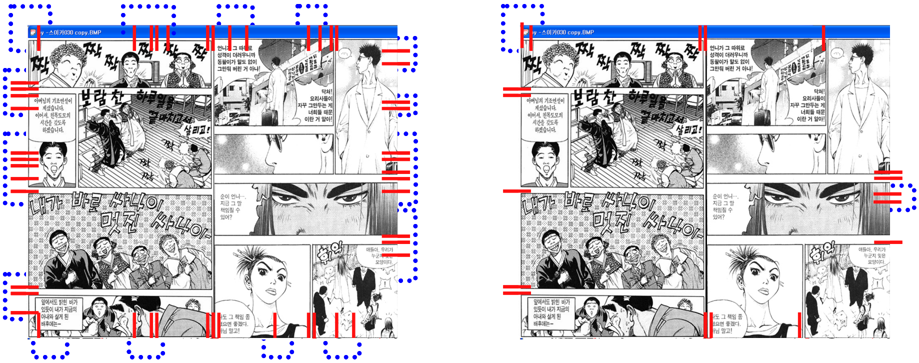


Fig. 6. The result of finding cutting-points



(a) (b)

Fig. 7. Errors on frame segmentation result. (a) the first forward image, (b) a result of bootstrap method (The box of the image boundary is errors.).

possible to segment artificially for off-line cartoon images by training images. Although the frame has some noises, the MLP can recognize the frame boundary. Then we can mark the cutting-point in the boundary and make the cutting-line. However, the result of the MLP is not complete.

As shown in Fig. 7, if the object is near the frame boundary, or the feature of the object in the frame is like a frame boundary, the MLP process can not recognize completely. To handle this problem we use the bootstrap method recommended by Sung and Poggio [18], which was initially developed to train neural networks for face detection. Some non-frame samples are collected and used for training. Plus, the partially trained MLP is repeatedly applied to images for more complete segmentation, and then patterns with a positive output are added to the training set as non-frame samples. This process iterates until no more patterns are added to the training set.

2.4 Verification

The output of MLP can have incorrect cutting-points. In the forward process, the cutting-line in the frame boundary shifts the position from top to bottom and from left

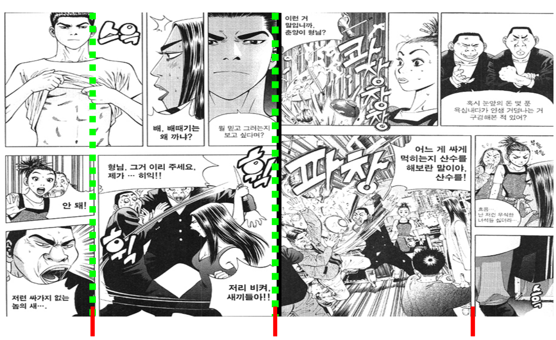


Fig. 8. One or more candidate cutting points

to right. If the input image has two cutting-points in the top and bottom, and that position is projected to the same position, the forward process in MLP also would make the cumulative cutting-point in the bottom and makes cutting-line. Then the candidate cutting-points are more than we would want to get. Fig. 8 shows this result. The dotted line of the image indicates the candidate cutting positions. Which one should we find the segmenting point? To handle this problem, we use the projection profile method. Each cutting-point, we can check the top-down pixels and find the real cutting-point which is the segmenting position that there is at least pixels in the axis of the position.

3 Experimental Results

This method is implemented in C++ language on an IBM-PC. 30 images of off-line cartoon were used to train MLP for the frame segmentation. The remaining 30 images were used for testing. Fig. 9 shows the frame segmentation results.

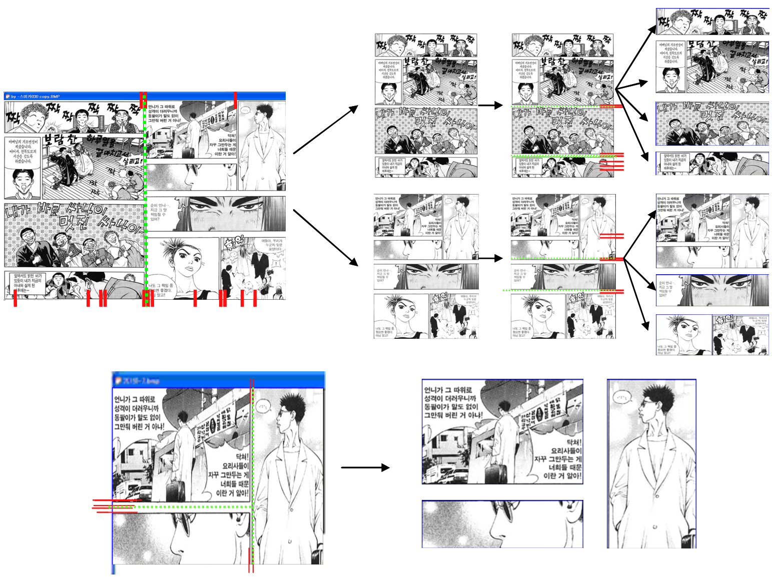


Fig. 9. One of the frame segmentation result for cartoon A images

The segmentation rates were evaluated using two metrics: precision and recall rate (Table 1). Equation (1) and (2) are the formula to compute the precision and recall rate. As shown in table 1, our method produced higher precision and recall rate than X-Y recursive cut algorithm without a MLP process, yet lower recall rates, as lack of training data for cartoons A, B and C.

$$precision\ (\%) = \frac{\#\,of\ \text{correctly detected cutting points}}{\#\,of\ \text{detected cutting points}} \times 100 \tag{1}$$

$$recall\ (\%) = \frac{\#\,of\ \text{correctly detected cutting points}}{\#\,of\ \text{desired cutting points}} \times 100 \tag{2}$$

Table 1. Comparison of precision rates

Cartoon Book Type	Category	without MLP process		with MLP process	
		Precision	Recall	Precision	Recall
Cartoon A	Training set	83.5%	78%	91.3%	96.5%
	Test set	-	-	87.7%	92.6%
Cartoon B	Training set	81.5%	76%	90.3%	92.5%
	Test set	-	-	87.3%	95.5%
Cartoon C	Training set	83.5%	78%	93.6%	98.5%
	Test set	-	-	87.2%	92.5%

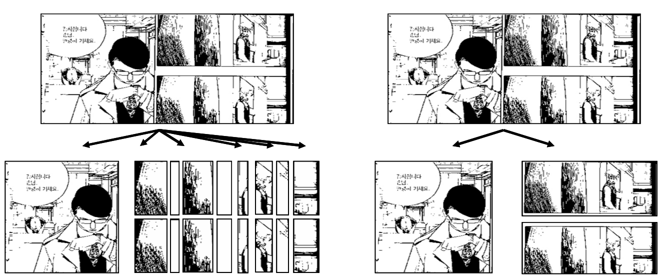
Table 2. Comparison of execution time

Measurement	with pre-process	without pre-process
Time (sec)	**2.8**	**10**

The execution time with pre-process is more efficient than without pre-process, because the size of input data from the cartoon is different (table 2). When we execute pre-process, we can reduce the input size.

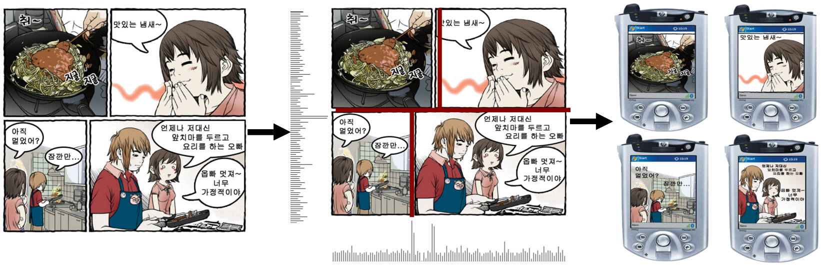
The segmentation errors in this experiment may classify into MLP-based segmentation step. This problem is mainly the result of a shortage of training data. An existing method for frame segmentation, X-Y recursive, has some problems because the noise and non-straight frame line affect the frame segmentation. The X-Y recursive method to compare the frame segmentation result is a sample data that the threshold value finding for the frame segmenting position is coordinated by our experimental environment. The results of this algorithm are not exact that show in Fig. 9 (a). Our new method using MLP-based X-Y recursive improves frame segmentation accuracy and variety of scanned images for transforming from off-line to mobile that Fig. 10 (b) shows this result.

Fig. 11 shows frame segmentation proposed result of mobile cartoon content that fit well on the mobile screen. The propose method has advantage to resize the cartoon content depends on the mobile device screen [17].



(a) (b)

Fig. 10. Comparing two methods. (a) a X-Y recursive result, (b) a MLP-based segmentation result.



(a) (b) (c)

Fig. 11. Frame Segmentation Proposes Result. (a) original image, (b) a MLP-based segmentation result, (c) mobile cartoon content.

4 Conclusion

Users generally access the cartoon content through mobile devices. This paper proposed a method for frame segmentation from a scanned image of paper-based cartoon image to the small screen mobile device**.** In this method, the segmentation process was implemented by a Multi-Layer Perceptron (MLP) trained by a back-propagation algorithm. The MLP-based frame segmentation process generates several candidate cutting points and the verification process using projection profile method recognizes these cutting-points in order to select the correct cutting point among candidates. Through experiments with various kinds of scanned images, it has been shown that the proposed method is very effective for the segmentation. However, various scanned images of off-line cartoon have a lot of frame features which are not a quadrangle. In this case, we can find the boundary point of segmenting frame, but our process can not segment inside the frame of the scanned images because of the non-quadrangle frame. As well, we are trying to segment the non-quadrangle frame and plan to extend our works to the frame that includes objects.

Acknowledgements. This work was supported by the Soongsil University Research Fund.

References

1. Wang, D., Srihari, S.N.: Classification of newspaper image blocks using texture analysis. Computer Vision Graphics, and Image Processing 47, 327–352 (1989)
2. Li, J., gray, R.M.: Text and picture segmentation by distribution analysis of wavelet coefficients. In: Proceedings of the 5th International conference on Image Processing Chicago, Illinois, pp. 790–794 (October 1998)
3. etemad, K., Doermann, D.S., Chellappa, R.: Multiscale segmentation of unstructured document pages using soft decision integration. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 92–96 (1997)
4. Pavlidis, T., Zhou, J.: Page segmentation by white streams. In: Proceedings of the First International Conference on Document Analysis and Recognition, St, Malo, France, pp. 945–953 (September 1991)

5. Akindele, O., Belaid, A.: Page Segmentation by Segment Tracing. In: proceedings of Second Int'l Conf. Document Analysis and Recognition, pp. 341–344 (1993)
6. Amamoto, N., Torigoe, S., Hirogaki, Y.: Block Segmentation and Text Area Extraction of Vertically/Horizontally Written Document. In: Proceedings of Second Int'l Conf. Document Analysis and Recognition, pp. 739–742 (1993)
7. Ittner, D., Baird, H.: Language-Free Layout Analysis. In: Proceedings of Second Int'l Conf. Document Analysis and Recognition, Tsukuba, Japan, pp. 336–340 (1993)
8. Antonacopoulos, A., Ritchings, R.: Flexible Page Segmentation Using the Background. In: Proceedings of 12 th Int'l Conf. Pattern Recognition, pp. 339–344 (1994)
9. Nagy, G., Seth, S.: Hierarchical Representation of Optically Scanned Documents. In: Proceedings of Seventh Int'l Conf. Pattern Recognition, pp. 347–349 (1984)
10. Krishnamoorthy, M., Nagy, G., Seth, S., Viswanathan, M.: Syntactic Segmentation and Labeling of Digitized Pages From Technical Journals. IEEE Trans. Pattern Analysis and Machine Intelligence 15, 743–747 (1993)
11. Pavlidis, T., Zhou, J.: Page Segmentation by White Streams. In: Proceedings of First Int'l Conf. Document Analysis and Recognition, pp. 945–953 (1991)
12. Ha, J., Haralick, R., Phillips, I.: Document Page Decomposition by the Bounding-Box Projection Technique. In: Proceedings of Third Int'l Conf. Document Analysis and Recognition, pp. 1119–1122 (1995)
13. Kise, K., Yanagida, O., Takamatsu, S.: Page Segmentation Based on Thinning of Background. In: Proceedings of 13th Int'l Conf. Pattern Recognition, pp. 788–792 (1996)
14. Fujisawa, H., Nakano, Y.: A Top-Down Approach for the Analysis of Documents. In: Proceedings of 10th Int'l Conf. Pattern Recognition, pp. 113–122 (1990)
15. Chenevoy, Y., Belaid, A.: Hypothesis Management for Structured Document Recognition. In: Proceedings of First Int'l Conf. Document Analysis and Recognition, pp. 121–129 (1991)
16. Ingold, R., Armangil, D.: A Top-Down Document Analysis Method for Logical Structure Recognition. In: Proceedings of First Int'l Conf. Document Analysis and Recognition, pp. 41–49 (1991)
17. Eunjung, H., Sungkuk, J., Anjin, P., Keechul, J.: Automatic Conversion System for Mobile Cartoon Contents. In: Proceedings of International Conference on Asian Digital Libraries, vol. 3815, pp. 416–423 (2005)
18. Sung, K.K., Poggio, T.: Example-based Learning for View-based Human Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1), 39–51 (1998)
19. Takehiko Inoue: SLAMDUNK. All rights reserved First published in Japan in, by Shueisha Inc., Tokyo Korean translation rights arranged with through Shueisha Inc. and Dai-Won Publishing Co., Ltd (1990)
20. Yim, J.O. : ZZANG. All rights reserved First published in Korea by DaiWon Inc.