DOI: 10.3745/KIPSTD.2010.17D.6.463

컬러 영상 위에서 DCT 기반의 빠른 문자 열 구간 분리 모델

신 현 경†

요 약

본 논문에서는 DCT 데이터에서 영상 데이터로의 해독 및 이진화 과정을 생략하고 컬러 영상의 DCT 관련 원자료를 사용하는 방법에 기반을 둔 매우 빠르고 안정적인 문자열 구간 분리 모형을 제안하였다.

DCT 블록에 저장된 DC 및 3개의 주요 AC 변수들을 조합하여 축소된 저해상도 회색 영상을 만들고 횡렬 및 종렬 투영법을 통해 얻어진 픽셀 값의 히스토그램을 분석하여 문자 열 구간 사이에 존재하는 백색의 띠 공간을 찾아내었다. 이 과정 중 탐색되지 않은 문자 열 구간은 마코프 모델을 사용하여 숨겨진 주기를 찾아내어 복원하였다.

본 논문에 실험 결과를 제시하였으며 기존의 방법보다 약 40 - 100배 빠른 방법임을 입증하였다.

키워드 : 문자열 구간 분리, 투영법, 비 부분 극한 제거, 마코프 모델링, 구역별 DCT(Discrete Cosine Transform), 영상 피라미드

Fast Text Line Segmentation Model Based on DCT for Color Image

Shin hyunkyung†

ABSTRACT

We presented a very fast and robust method of text line segmentation based on the DCT blocks of color image without decompression and binary transformation processes.

Using DC and another three primary AC coefficients from block DCT we created a gray-scale image having reduced size by 8x8. In order to detect and locate white strips between text lines we analyzed horizontal and vertical projection profiles of the image and we applied a direct markov model to recover the missing white strips by estimating hidden periodicity.

We presented performance results. The results showed that our method was 40 - 100 times faster than traditional method.

Keywords : Text Line Segmentation, Projection Profile Method, Non-maximal Suppression, Markov Modeling, Block DCT (Discrete Cosine Transform), Image Pyramid

1. Introduction

OCR performance is greatly degraded when layout structure of document is complicated, e.g. invoice documents with forms, magazine pages with pictures embedded. Text line is by definition a linear or smooth region consisting of rectangular regions containing word. Since merging text lines into paragraph is an essential part of document layout process, successful acquisition of text line naturally guides OCR engines to achieve reliable text regions consisting of well behaved text paragraphs. Standard OCR engines internally incorporate document layout segmentation process in order to enhance overall OCR performance.

In text documents recognition, line segmentation is a modeling problem of three variables [1]: skew variability, line proximity, and fragmentation. In this paper, we ignore skew variablity to focus on document image acquired from calibrated devices. Major categories of line segmentation algorithms are as follows: projection profile

※ 본 연구는 2010년 경원대학교 교내 연구비 지원을 받아 시행된 연구임.
† 종신회원 : 경원대학교 수학정보학과 조교수
논문접수 : 2010년 6월 21일
수 정 일 : 1차 2010년 8월 5일
심사완료 : 2010년 8월 6일

with CTM (cut text minimization) or XY cut; RLSA (run length smearing algorithm) or fuzzy RLSA; grouping; Hough transform; repulsive attractive network; stroke filter; and HMM (hidden markov model). Further details are elaborated in the next section.

Most of approaches mentioned above are based on binary image. In this paper we present a method based on gray scale image with a statistical modeling framework. Additionally, our method uses block DCT to construct reduced size low resolution gray images. The method with low resolution is for speed up and statistical modeling for high accuracy.

This paper organizes as follows: in section 2 related research works are introduced, in section 3 text line segmentation algorithms are explained, in section 4 experimental results are presented, and in section 5 future works on this method are discussed.

2. Related Works

Projection based method uses projection profile of pixel values along with specific directions. Basic recursive XY mincut on binary image is introduced in [2], which is used widely. Haralick [3] suggested horizontal and vertical projection on the bounding boxes of connected components rather than binary image itself. Bounding box of a connected component is filled as black. This method improves the traditional projection method with high cost of computational complexity to retrieve connected components. To relieve the complexity, in [4], projection profile on edge map is considered. But the reduced complexity is not enough. The weakest aspect of projection profiling is its directional dependency, i.e. poor results from the projection on skewed document. To minimize the directional dependency, piecewise projection is employed by [5] in which document is divided into 8 columns. Unlike the previous methods based on binary image, projection profile on gray scale is studied in [6]. This method implements gaussian convolution and automatic detection of scale space which requires large computational complexity.

Smearing method is bottom-up method based on run-length smoothing algorithm(RLSA). The basic idea is introduced in [2]. Focusing on separation of overlapping and touching text strokes for which standard RLSA is not effective, fuzzy run-length smoothing algorithm is suggested in [7]. However the performance is still sensitive on the choice of parameter values defining fuzzy algorithm and on the choice of 1D or 2D smoothing scheme.

Grouping method is representative of bottom-up approach using feature unit in document such as pixel, connected components, or salient points. Among several approaches, baseline segment [8], edge map on variance image [9], and level set [10] are considered as salient feature set, respectively. As a common weakness of bottom-up method, computational complexity is high for all of those methods mentioned above.

As a classical and robust method introduced in [2], hough transformation can be applied to fluctuating lines. However the method is not useful with presence of graphics in document.

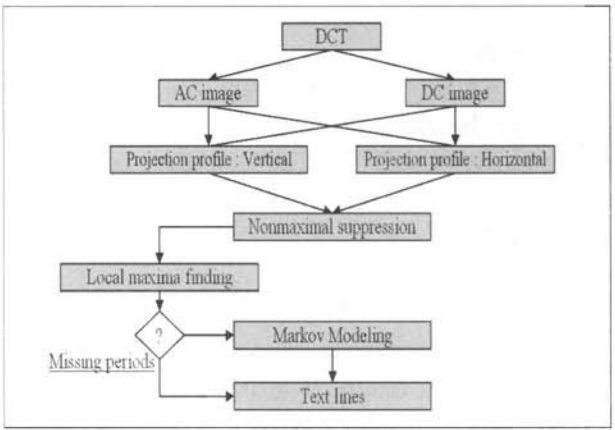
Machine learning theory is a good candidate to recognize line patterns. Oztop et al. [11] adopt repulsive attractive network to locate baseline in document as a implementation of self-organization map framework. New clustering technique of distance learning is recently applied to line segmentation problem in [12].

As non-traditional other methods, stochastic method of hidden markov model [13], stroke filter method based on morphology operation [14], wavelet analysis [15] are published.

In this paper, we apply classical projection method on hierarchical image pyramid retrieved from DCT block. Simplicity of projection method and reduced size image space in pyramid greatly improve the speed. Performance is improved through markovian modeling of periodicity of line location.

3. Text Line Segmentation Model

As a part of project for real-time document processing, our method is applied on the 8x8 size reduced low resolution image corresponding to the third level in image pyramid. To retrieve 8x8 reduced low resolution image and its edge map, we access to DCT block in JPEG stream. There is almost no processing time to get the two sub-images. Projection profile method is employed on the two sub-images to locate the horizontal white strips and the vertical columns. For the horizontal white strips, projection profile on AC image is considered and for the vertical column strips, projection profile on DC image is



(Fig. 1) Overview of line segmentation

considered.

3.1 Construction of Reduced Size Gray Scale Images Using DCT

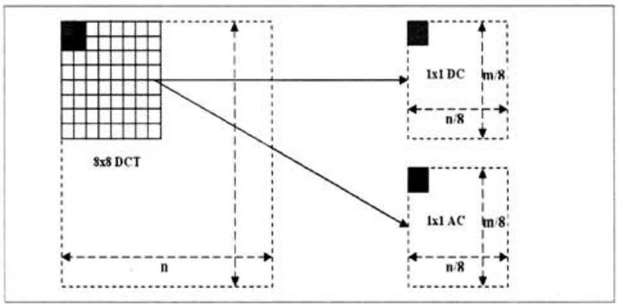
Given a JPEG image, without decompression of DCT encoded image data, raw DCT data can be retrieved. For the convenience of notation, let's denote 8x8 DCT block matrix as below in (Fig. 2):

a_{00}	a_{01}	a_{02}	a_{03}	a_{04}	a_{05}	a_{06}	a_{07}
a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}
a_{20}	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}
a_{30}	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}
a_{40}	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}
a_{50}	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}	a_{56}	a_{57}
a_{60}	a_{61}	a_{62}	a_{63}	a_{64}	a_{65}	a_{66}	a_{67}
a_{70}	a_{71}	a_{72}	a_{73}	a_{74}	a_{75}	a_{76}	a_{77}

(Fig. 2) matrix representation of DCT block

Using the block DCT we construct two gray scale images, AC composition image and DC image, as the following way. For construction of AC image, The three primary AC coefficients (a_{01}, a_{10}, a_{11}) are averaged to represent the corresponding pixel value. For construction of DC image, DC coefficient (a_{00}) is selected to represent the corresponding pixel value. (Fig. 3) shows the process.

AC image corresponds approximately to edge strength of level three sub-image in image pyramid while DC image to average value of corresponding block.

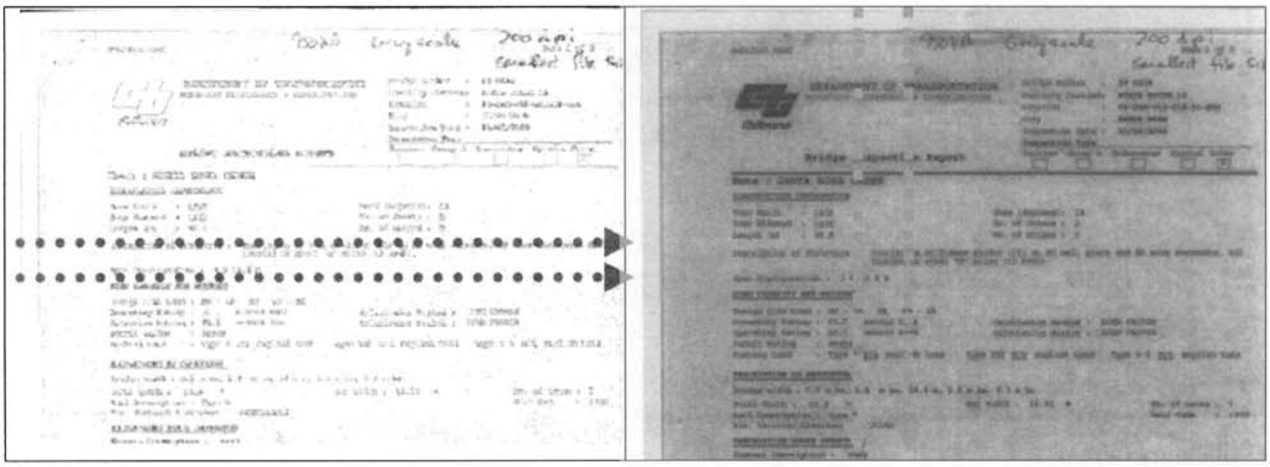


(Fig. 3) Construction of AC and DC images from block DCT

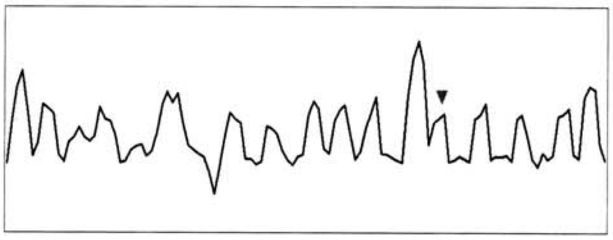
3.2 Projection Profile and Min Cut for Text Line Segmentation

In order to find white strips between text lines, projection profile method is employed on the AC image constructed at the previous section. As seen in (Fig. 4) vertical projection profile, h(y), is built as follows: for each y position ranging from the top to the bottom, sum of pixel values are estimated along with horizontal direction. For horizontal projection profile, v(x), at each x position ranging from the left to the right sum of the pixel values are estimated along with vertical direction.

(Fig. 5) shows the resulting vertical projection profile, h(y), on the sample image in (Fig. 4). The maximal



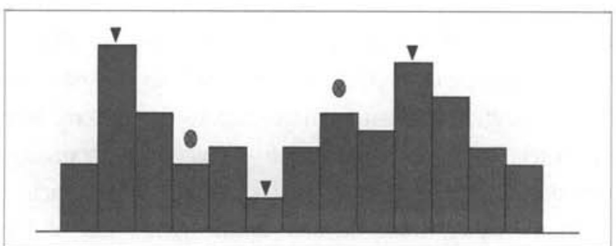
(Fig. 4) schematic visualization of horizontal projection on AC image and vertical projection on DC image



(Fig. 5) graphical representation of vertical projection profile of the sample image

points of h(y) indicate strong evidence of presence of white strips.

Our local maxima finding algorithm is based on non-maximal suppression: after suppress the non-maximal and minimal points within window size of 3, we filter out weak gradient point. Among strong gradient points, we match the pair of maximal and minimal points. Some of the maximal and minimal points are further filtered out if there is no bijective matching of alternative occurrence. As seen in (Fig. 6), non-maximal supression leaves 5 extreme points (blue and red). Our min-cut algorithm of bijectivity can filter out two spurious maximal points (colored red).



(Fig. 6) illustration of finding local extreme points. The blue triangles indicate true local extremes while red circle do noise in profile.

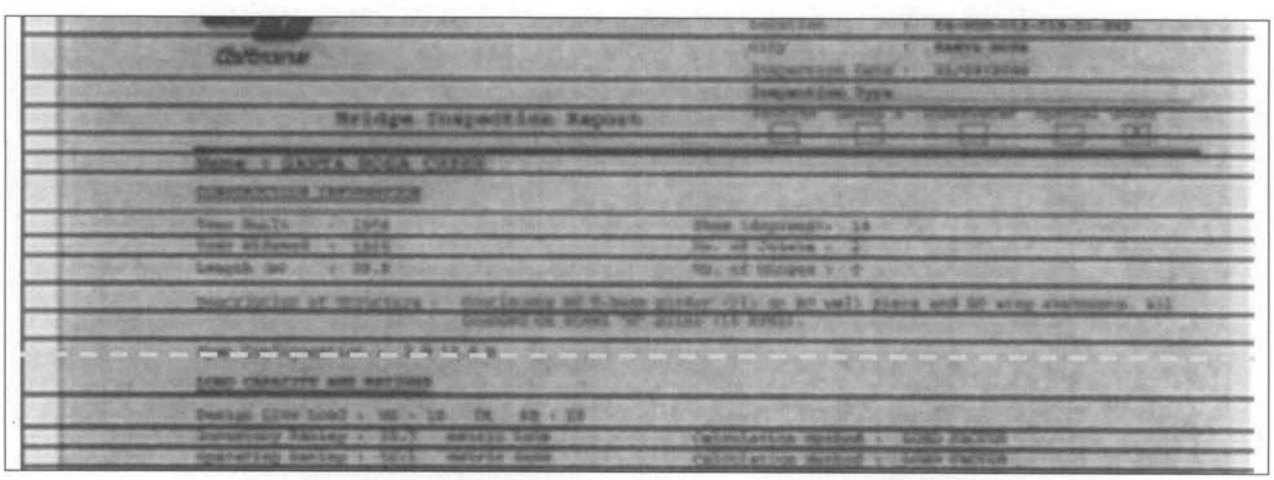
A result of min cut is displayed in (Fig. 7). The red lines in (Fig. 7) correspond to local maxima in (Fig. 5). From the bottom, there is a missing white strip (yellow dotted line) between the fourth and the fifth. The point under the red triangle in (Fig. 5) is correspond to the missing strip. To resolve this problem, we employed markov modeling.

3.3 Markov Model For Dectecing Missing Text Lines

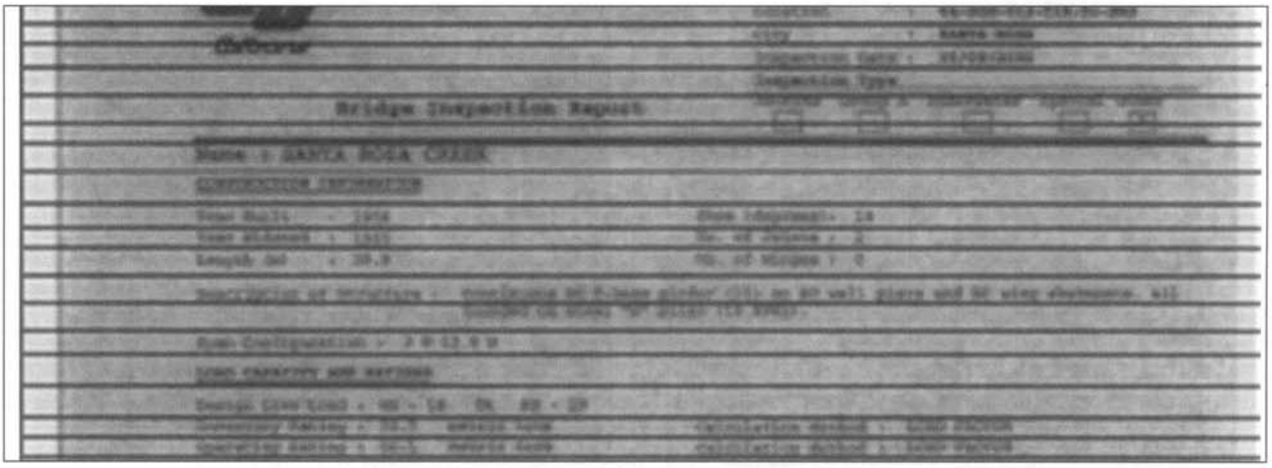
As seen in (Fig. 7), projection profile method misses white strips when density of texts are not uniform over the text lines. We employ markov modeling to recover the missing features. We denote a markov model M = M(w, P), where w and P denote the finite states and the transition probability matrix, respectively. The random variable w represent height of text line, ranging from 1 to n. For the convenience, we set n = 16.

For a given text line, heights of upper and lower adjacent text lines are considered as random variable. We design two step markov process which possesses two stage history values. construction of transition matrix is as follows: p_{ijk} is probability of current text line being w_i while heights of previous and next text lines being w_j and w_k. Intuitively, it is very high probable for $w_i = w_j$ if w_j and w_k are the same. Likelihood function is set as $p(x|w_i) = G(x, w_k, w_j, \sigma=16)$, where G denotes gaussian mixture with two means of w_j and w_k and single variance σ. Using the likelihood, $p(w_i|x)$ evaluates the confidence level of presence of hidden period between two text lines.

In (Fig. 8) we illustrate that a missing text line is recovered throuhg markov modeling (see the fifth text line from the bottom).



(Fig. 7) illustration of the white strips obtained from min-cut algorithm without consideration of periodicity of occurrecne



(Fig. 8) recovery of missing text line using Markov Model

4. Experimental Results

4.1 Data

As seen in <Table 1> we used 5 different type of document images. The documents were obtained from the scanned document archives in the department of information mathematics.

For the study of this paper, we used 1784 simple text images (TYPE 1 - TYPE 3) and 504 complicated color document images (TYPE 4 - TYPE 5). We visualized the text line segmentation results from sample images in (Fig. 9), in which type 1 (left-top), type 2(right-top), type 3(left-mid), type 4 (right-mid), and type 5 (left and right bottom) were placed. Text line segmentation performed very well on all types of document images. For convenience of visualization of the colored text lines, some of the color images were converted to gray scale.

4.2 Test Results

The simulation has been performed on the single core processor with 1.90 GHz CPU speed and 4 GB memory space.

<Table 2> shows the processing times for each catergory of dataset as described in <Table 1>. It displays in each document category as well as the minimum, the first quatile, the median value processing

〈Table 1〉 category of image data used in this paper

Document Type	Number of Files	Size in Pixels	Color	Description
TYPE 1	241	2488x3222	Gray	Clean and full text document
TYPE 2	509	2544x3392	Gray	Full text document
TYPE 3	1,034	2544x3298	Gray	Texts and graphics with simple layout structure
TYPE 4	99	2552x3300	Full color	Complicated layout
TYPE 5	405	2400x3300	Full color	Low contrast with complicated layout structures

〈Table 2〉 processing time (msec x 1000) table for 5 different datasets: DCT method (upper) and traditional method (lower)

	MIN	Q1	MEDIAN	Q3	MAX
TYPE 1	14,223	15,062	15,474	16,314	532,736
	97,273	717,372	1,091,705	1,535,866	3,115,429
TYPE 2	14,670	17,577	18,686	20,585	545,328
	96,112	763,353	1,122,569	1,482,840	2,696,787
TYPE 3	13,596	16,266	13,596	18,556	566,851
	30,189	1,081,085	1,286,324	1,615,657	4,020,915
TYPE 4	1,571	8,829	18,928	23,019	612,694
	72,450	502,131	964,931	1,500,216	6,719,430
TYPE 5	5,407	16,812	18,531	20,108	586,070
	9,582	748,859	1,168,933	1,606,712	3,546,052

(Fig. 9) illustration of text line segmentation on sample images from the five types of data sets: type-1, type-2, type-3, type-4, type5, type-5, from top to bottom and from left to right.

time, the third quartile, and the maximum processing time, from left to right in column, respectively. For each type of the document, our DCT method is shown the upper row and the conventional method (using the projection profile method[2]) is shown in the lower row. The time scale is 1000 times of a milli-second.

<Table 3> shows the ratio of processing times between the two methods. MIN column indicates the best case,

	MIN	Q1	MEDIAN	Q3	MAX
TYPE 1	6.83	47.62	70.55	94.14	5.85
TYPE 2	6.55	43.42	60.08	72.03	4.95
TYPE 3	2.22	66.46	75.76	87.07	7.09
TYPE 4	46.12	56.87	50.98	65.17	10.97
TYPE 5	1.77	44.54	63.08	79.90	6.05

〈Table 3〉 ratios of processing times (conventional method / DCT method)

i.e., the samples taking short processing time. In the case MIN column, our method improved from 1.77 to 46.12 times faster. MAX column indicates the worst case, .e., the samples taking long processing time. In such case, our method is faster from 4.95 to 10.97 times. However, huge improvement was observed at the inter-quartile range and the median cases. Our method was faster in as much as from 43 to 97 times) than the conventional method.

5. Remarks

The document segmentation is a very hard subject if it is not unsolvable. The document layout analysis is a part of document segmentation. The text line segmentation is a basic part of the document layout analysis.

In this paper we presented a very fast method for text line segmentation. The speed came from that our method skipped the stage of decompressing the image data. We will use the DCT based text segmentation method as the base for the more complicated layout analysis.

The output text of OCR is usually unstructured. The conventional text processing methods, using the query and correction, are good for the well structured text data (such as web page). The text line segmentation transfers the unstructured texts to the semi-structured texts, which results in better use of the text processing methods.

When it comes to perform OCR with the color documents, the accuracy is usually poor compared with the case of the binary documents. For this reason, most of the OCR engines take the binary image as the input, which makes the binary conversion is unavoidable for OCR with the color image. However, a good binary conversion takes a long processing time. In this context, a fast and accurate binary conversion is important. Our DCT method also has potential to be extended to binarization.

References

[1] L.L. Sulem, A. Zahour, B. Taconet, "Text Line Segmentation of Historical Documents: a Survey", IJDAR 2007.

[2] Managing Gigabytes

[3] Ha, J., Haralick, R. and Phillips, I., "Recursive X-Y Cut Using Bounding Boxes of Connected Components", Proc. Third Int'l Conf. Document Analysis and Recognition, pp.952-955, 1995.

[4] A. Zahour, B. Taconet, P. Mercy, and S. Ramdane, "Arabic hand-written text-line extraction", ICDAR 2001.

[5] R. Ryue, J. Song, M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction", IEEE Trans. CSVT, 2005.

[6] R. Manmatha, N. Srimal, "Scale space technique for word segmentation in handwritten manuscripts", PAMI, 2005.

[7] Shi, Z., Venu Govindaraju, "Line separation for complex document images using fuzzy runlength", Proceedings. First International Workshop, 2004.

[8] M. Feldback, K.D. Tonnies, "Line Detection and Segmentation in Historical Church Registers", ICDAR, 2001.

[9] Y. Zhong, K. Karu, A.J. Jain, "Locating Text in Complex Color Images", PR, 1995.

[10] Y. Li, Y. Zheng, D. Doermann, "Script-independent Text Line Segmentation in Freestyle Handwritten Documents", IEEE Trans. PAMI., 2008.

[11] E. Oztop et al, "Repulsive attractive network for baseline extraction on document Images", IEEE Signal proceesing. 1997.

[12] F. Yin, C. Liu, "Handwritten Chinese text line segmentation by clustering with distance metric learning", PR, 2009.

[13] Tseng, Lee, "Recognition based handwritten Chinese character segmentation using a probabilistic Viterbi algorithm", PR Letter, 1999.

[14] C. Jung, Q. Liu, J. Kim, "A new approach for text segmentation using a stroke filter", Signal Processing, 2008.

[15] Q. Ye, Q. Huang, W. Gao, D. Zhao, "Fast and robust detection in images and video frames", Image and Vision Computing, 2005.

신 현 경

e-mail : hyunkyung@kyungwon.ac.kr

2002년 State University of New York at Stony Brook. 응용수학과(공학박사)

2008년~현 재 경원대학교 수학정보학과 조교수

관심분야 : Image Processing. Neural Network. Machine Learning.