

MSER을 이용한 문서 이미지 이진화 기법

유영중*

Document Image Binarization Technique using MSER

Young-jung Yu*

Department of Computer Engineering, Pusan University of Foreign Studies, Pusan 609-815, Korea

요 약

문서 이미지의 이진화는 문서 인식의 이전 단계에서 주로 사용되며, 이진화의 성공 여부에 따라 문서 인식의 결과에 영향을 미치는 중요한 단계로 볼 수 있다. 지금까지 문서 이미지를 이진화 하기 위한 다양한 기법들이 연구되었지만, 문서 이미지의 상태에 따라 그 결과는 다양하다. 본 논문에서는 객체 추출에 많이 이용되는 MSER(Maximally Stable Extremal Region)을 이용하여 문서 이미지를 이진화하는 기법을 제안한다. 먼저 문서 이미지에서 MSER 객체를 추출한다. 추출된 MSER 객체는 그 자체로 문서 이미지 이진화에 사용되기는 어렵기 때문에 사용하기 적합한 형태로 변경되는 과정을 거친다. 그리고 최종 MSER 객체와 문서 이미지로부터 추출한 대비 이진 이미지를 이용하여 최종 이진 이미지를 계산한다. 실험결과는 본 논문에서 제안한 방법이 문서 이미지의 이진화에 유용함을 보여준다.

ABSTRACT

Document image binarization is largely used as previous stage of document recognition. And the result of document recognition is much affected from the result of document image binarization. There were many studies to binarize document images. The results of previous studies for document image binarization is varied according to the state of document images. In this paper, we propose a technique for document image binarization using MSER that is applied to extract objects from an image. At first, raw MSER objects are extracted from a document image. Because the raw MSER objects cannot be used for document image binarization, the extracted raw MSER objects are modified. Then the final MSER objects are used for document image binarization with the contrast image that is extracted from the document image. Experimental results show that the proposed technique is useful for document image binarization.

키워드 : MSER, 문서 이미지 이진화, 대비 이미지, 문서 인식

Key word : MSER, Document Image Binarization, Contrast Image, Document Recognition

접수일자 : 2014. 04. 24 심사완료일자 : 2014. 06. 03 게재확정일자 : 2014. 06. 16

* Corresponding Author Young-jung Yu(E-mail:yjyu@bufs.ac.kr, Tel:+82-51-509-6224)

Department of Computer Engineering, Pusan University of Foreign Studies, Pusan 609-815, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2014.18.8.1941>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

일반적으로 하나의 문서는 배경과 텍스트만으로 구성이 되거나 배경, 텍스트, 그리고 그림으로 구성된다. 본 논문의 대상은 그림을 제외한 텍스트와 배경으로 구성된 문서 이미지이다. 텍스트와 배경으로 구성된 문서 이미지로부터 텍스트를 인식하고 자동으로 전자문서를 생성하는 것은 이미 일반화된 작업이다. 문서 이미지로부터 텍스트를 인식하기 위해서는 문서 이미지에서 배경과 텍스트를 분리하는 작업이 선행되어야 한다. 이것을 위해서는 임의의 픽셀이 배경과 텍스트 중 어느 부분에 해당하는지를 결정해야한다. 임의의 픽셀이 배경과 텍스트 중 어느 부분에 해당하는지를 결정하기 위해 필요한 작업이 문서 이미지의 이진화이다.

문서 이미지의 이진화를 위한 입력 이미지는 회색 이미지를 사용한다. 문서 이미지의 이진화를 위한 방법 중 가장 먼저 고려해 볼 수 있는 방법이 전역 문턱치(global threshold) 값을 사용하는 방법이다. 입력 회색 이미지의 밝기 값이 이미지 전체에서 일정한 경우 하나의 전역 문턱치 값 만으로 텍스트와 배경을 분리할 수 있다. 그러나 문서 이미지마다 배경과 텍스트를 분리하는 전역 문턱치 값이 상이할 수 있기때문에, 입력 이미지에 적합한 전역 문턱치 값을 자동으로 계산하는 방법이 필요하며 Otsu[1]에 의해 제안되었다. Otsu[1]가 제안한 방법은 문서 이미지의 밝기가 일정한 경우에 문제없이 작동하지만, 밝기가 일정하지 않은 문서에서는 텍스트와 배경을 분리하는데 어려움이 있다. 이 문제를 해결하기 위해 제안된 방법이 지역 문턱치(local threshold) 값을 이용하는 방법이다. 이미지의 각 픽셀을 중심으로하는 지역 창(local window)을 만들고, 각 픽셀에 적합한 문턱치 값을 계산한다. Niblack[2]은 평균과 표준편차를 이용하여 식(1)과 같이 문턱치 값을 계산하는 방법을 제안하였다.

$$T = m + k \times s \quad (1)$$

위 식에서 m 은 평균, s 는 표준편차이고 k 는 입력 가능한 상수로 -0.2 를 사용하였다. Sauvola등[3]은 Niblack[2]의 방법을 개선하여 좀 더 나은 이진화 결과를 만들 수 있는 방법을 식(2)와 같은 방법으로 제안하였다.

$$T = m \times \left(1 - k \times \left(1 - \frac{s}{R} \right) \right) \quad (2)$$

위 식에서 m 은 평균, s 는 표준편차, k 는 임의의 상수 값으로 0.5 가 사용되었고, R 은 최대표준편차 값으로 128 이 사용되었다. Niblack[2]과 Sauvola[3] 외에도 다양한 방법으로 지역 문턱치 값을 계산하는 방법들이 제안되었다. 지역 문턱치 값을 이용하는 방법들이 전역 문턱치 값을 이용하는 방법에 비해 좋은 결과를 보이지만, 이미지내의 픽셀 값들의 대비에 민감하고 지역 창 크기에 따라 결과가 상이할 수 있는 단점이 있다.

Gatos등[4]은 이전과는 다른 방법을 제안했다. 먼저 Sauvola등[3]의 방법을 이용해 대략적으로 텍스트 영역과 배경 영역을 구분하고, 배경 영역에 해당하는 픽셀들을 이용해 텍스트 영역의 픽셀들에 대한 밝기 값을 추정한다. 다음으로 텍스트 영역의 픽셀들에 대해 추정된 밝기 값과 원래 밝기 값을 비교하여 텍스트 픽셀 여부를 결정한다. Gatos등[4]의 방법에서는 초기에 Sauvola등[3]의 방법에 의해 대략적으로 추정되는 텍스트 영역이 실제 텍스트에 포함되는 픽셀들을 포함하지 않을 경우 좋은 결과를 기대하기 어렵다. Lu등[5]은 샘플링을 통해 배경 픽셀 값들을 추정하고, 추정된 배경 픽셀 값들을 이용해 텍스트의 경계 픽셀들을 추출한다. 추출된 경계 픽셀 값들을 이용해 텍스트의 폭을 추정하고, 이를 기반으로 각 픽셀의 텍스트 여부를 판정한다. Lu등[5]의 방법에서의 관건은 추정된 배경을 기반으로한 경계 픽셀들의 추정의 정확성과 텍스트 폭 추정의 정확성이다. 하나의 이미지 내에 폭이 다른 여러 종류의 텍스트가 포함된 경우 좋은 결과를 얻기가 힘들다. Su등[6,7]은 대비 이미지를 통한 대략적인 텍스트 픽셀 추정과 텍스트 폭 추정을 통해 문서 이미지를 이진화하는 방법을 제안하였다. Su등[6,7]의 방법도 초기에 추정되는 텍스트 픽셀의 정확도가 낮은 경우 만족할만한 결과를 보여주지 어렵다. 또한 다양한 텍스트 폭을 가지는 문서 이미지에는 적용이 어려운 단점이 있다.

본 논문에서는 Su등[6,7]이 사용한 대비 이진 이미지와 Matas등[8]이 제안한 MSER을 이용하여 문서 이미지를 이진화하는 방법을 제안한다. II장에서는 Su등[6,7]이 제안한 대비 이미지에 대해 간단히 설명하고,

Matas등[8]이 제안한 MSER을 본 연구에 적합한 형태로 변경하는 방법과 마지막으로 문서 이미지를 이진화하는 방법에 대해 설명한다. III장에서는 다양한 문서 이미지에 대한 실험결과를 소개하고 IV장에서 결론을 맺는다.

II. MSER을 이용한 이미지 이진화

본 논문에서는 입력 문서 이미지 I_{in} 의 이진화를 위해 대비 이진 이미지 I_{cb} 와 MSER 이미지 I_{MSER} 를 사용한다. 대비 이진 이미지에는 추출되어야 할 텍스트 픽셀들의 부분집합(subset)이 포함되어 있는 것으로 가정하고, MSER 이미지는 확대집합(superset)이 포함되어 있는 것으로 가정한다. 최종적으로 두 결과를 이용하여 이진화된 문서 이미지 I_{binary} 를 생성한다.

2.1. 대비 이진 이미지

Su등[6]은 이미지에 포함된 텍스트의 경계 픽셀을 얻기 위해 식(3)을 이용하였다.

$$I_c(x,y) = \frac{f_{\max}(x,y) - f_{\min}(x,y)}{f_{\max}(x,y) + f_{\min}(x,y) + \epsilon} \quad (3)$$

위 식에서 $f_{\max}(x,y)$ 와 $f_{\min}(x,y)$ 는 픽셀 (x,y) 를 중심으로하는 지역 창에서 가장 큰 밝기 값과 가장 낮은 밝기 값을 의미한다. 지역 창의 크기는 3×3 이 사용되었다. 식(3)을 통해 얻은 각 픽셀의 대비 값들을 기반으로 텍스트와 배경 픽셀 영역을 분리하기 위해 이미지 I_c 에 Otsu[1]의 방법을 적용하여 대비 이진 이미지 I_{cb} 를 얻을 수 있다. 그림 1은 입력 이미지와 식(3)을 적용하여 얻은 대비 이미지 그리고 대비 이미지에 Otsu[1] 방법을 적용한 대비 이진 이미지를 보여주고 있다. 그림 1에서 보는 바와 같이 대비 이진 이미지는 약간의 잡음을 포함하지만 대부분 텍스트 영역에 해당하는 픽셀들의 부분집합으로 구성되어 있음을 알 수 있다.

2.2. MSER 이미지

MSER은 Matas등[8]이 이미지로부터 관심 영역을 추출하기위해 제안한 방법이다. Matas등[8]이 제안한 MSER 추출 방법을 회색 문서 이미지에 적용할 경우 텍스트 영역에 해당하는 대부분의 픽셀들과 일부 배경 영역에 해당하는 픽셀들이 추출된다.



그림 1. 대비 이미지와 대비 이진 이미지: (a) 입력 이미지, (b) 대비 이미지, (c) 대비 이진 이미지

Fig. 1 Contrast image and contrast binary image: (a) input image, (b) contrast image, (c) contrast binary image

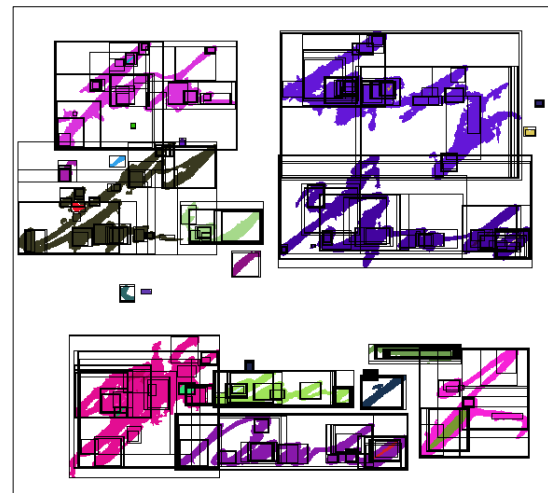


그림 2. MSER 이미지

Fig. 2 MSER image

그림 2는 그림 1의 입력 이미지에 MSER을 적용한 결과 그림이다. 그림에서 흰색은 배경 픽셀이고 나머지 픽셀들이 텍스트 후보 픽셀들이다. 그림 1의 입력 이미지와 비교해볼 때 그림 2의 텍스트 후보 픽셀들이 그림 1의 실제 텍스트 픽셀들을 대부분 포함하고 있음을 볼 수 있다. 그림 2에서 사각형 영역은 MSER의 결과로 추출된 영역들이다. 그림 2에서 보는 바와 같이 많은 작은 사각형 영역들이 큰 사각형 영역의 내부에 포함되어있다. 본 논문에서는 텍스트 후보 픽셀들을 포함하는 하나의 사각형 영역만을 필요로하기 때문에 다른 사각형 영역내에 포함되어 있는 사각형 영역을 제거한다.

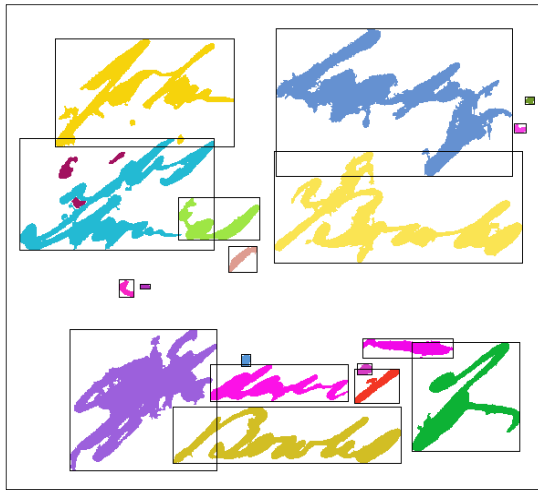


그림 3. 내부 사각형들이 제거된 MSER 이미지
Fig. 3 MSER image of which inner rectangles are eliminated

그림 3은 그림 2의 MSER 이미지에서 다른 사각형 영역에 포함되는 내부 사각형 영역들을 제거한 후의 결과이다. 그림 1의 입력 이미지와 비교해보면 일부 배경 픽셀에 해당하는 작은 사각형 영역이 있음을 확인할 수 있다. 배경 픽셀들만을 포함하는 작은 사각형 영역을 제거하기 위해 이미지 I_{cb} 를 이용하여 식(4)와 같이 계산한다.

$$S = \{(x, y) | I_{cb}(x, y) = 0, (x, y) \in R_i\} \quad (4)$$

$$R_i \text{ is removed, if } \frac{N_S}{N_{R_i}} < k$$

위 식에서 R_i 는 그림 3의 단계에서의 i 번째 사각형 영역안의 텍스트 후보 픽셀들을 의미하고, N_S 는 집합 S 의 픽셀 수, N_{R_i} 는 R_i 의 픽셀 수를 의미한다. k 는 상수이며 본 논문에서는 $k = 0.1$ 을 사용한다. 식(4)를 그림 3의 단계에 적용하면 배경 픽셀들로 구성된 사각형 영역들을 제거할 수 있다.

그림 4는 그림 3의 단계에 식(4)를 적용한 결과이다. 그림 3에서 배경 픽셀들만을 포함하고 있는 4개의 사각형 영역이 제거되었음을 확인할 수 있다.

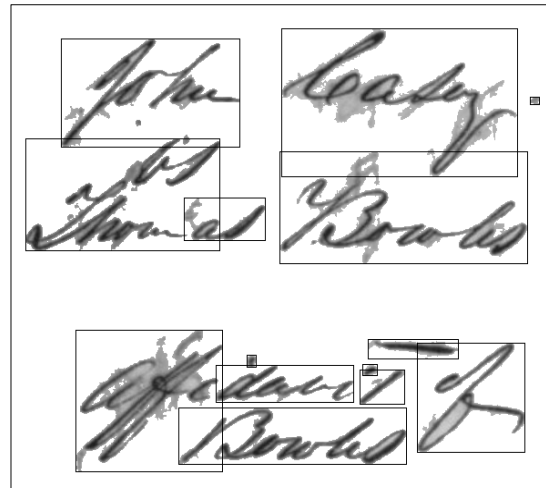


그림 4. 최종 MSER 이미지
Fig. 4 Final MSER image

2.3. 이진화

그림 4와 같은 최종 MSER 이미지에서의 배경 픽셀들은 I_{binary} 에서 배경 픽셀에 해당한다. I_{MSER} 에서 배경 픽셀이 아닌 나머지 픽셀들은 I_{in} , I_{cb} 그리고 I_{MSER} 을 이용하여 다음과 같이 계산된다.

$$Q = \{I_{in}(x, y) | I_{cb}(x, y) = 0, (x, y) \in R_i\} \quad (5)$$

$$I_{binary}(x, y) = \begin{cases} 0 & I_{in}(x, y) \leq Q_{mean} + Q_{std}/2 \\ 1 & otherwise \end{cases}$$

위 식에서 R_i 는 i 번째 사각형 영역안의 텍스트 후보 픽셀들을 의미하고, Q_{mean} 은 집합 Q 의 픽셀 값들의 평균, Q_{std} 는 표준편차를 의미한다. 그림 5는 식(5)를 적용

한 이진화 결과 이미지를 보여준다.

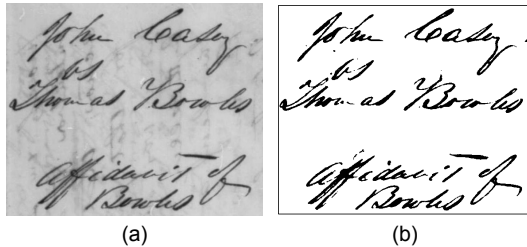


그림 5. 이진화 결과: (a) 입력 이미지, (b) 결과 이미지

Fig. 5 Binarization result: (a) input image, (b) result image

III. 실험결과

본 논문에서는 실험을 위해 OpenCV 2.1에서 제공하는 MSER 관련 함수를 사용하고, DIBCO에서 제공하는 이미지들을 테스트 이미지로 사용하였다.

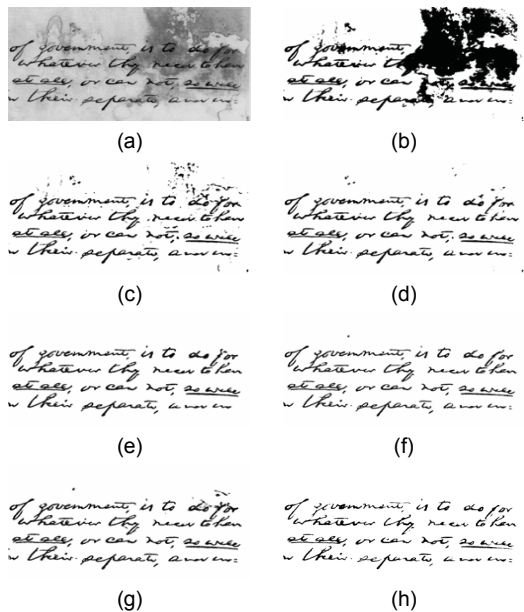


그림 6. 이진화 결과들의 비교 (a) 입력 이미지 (b) Otsu[1] (c) Sauvola[3] (d) Gato[4] (e) Lu[5] (f) Su[6] (g) Su[7] (h) Our result

Fig. 6 Compare of binarization results (a) input image (b) Otsu[1] (c) Sauvola[3] (d) Gato[4] (e) Lu[5] (f) Su[6] (g) Su[7] (h) Our result

그림 6은 저하된 이미지에 대한 각 이진화 방법들의 결과를 보여준다. 전역 문턱치나 지역 문턱치를 이용한 (b), (c)의 경우 좋은 결과를 얻기 어려움을 알 수 있다. Lu[5]와 Su[6,7]의 결과가 본 논문의 결과와 유사하지만, 그들의 방법은 전체 텍스트 폭이 일정해야 한다는 단점을 가진다.

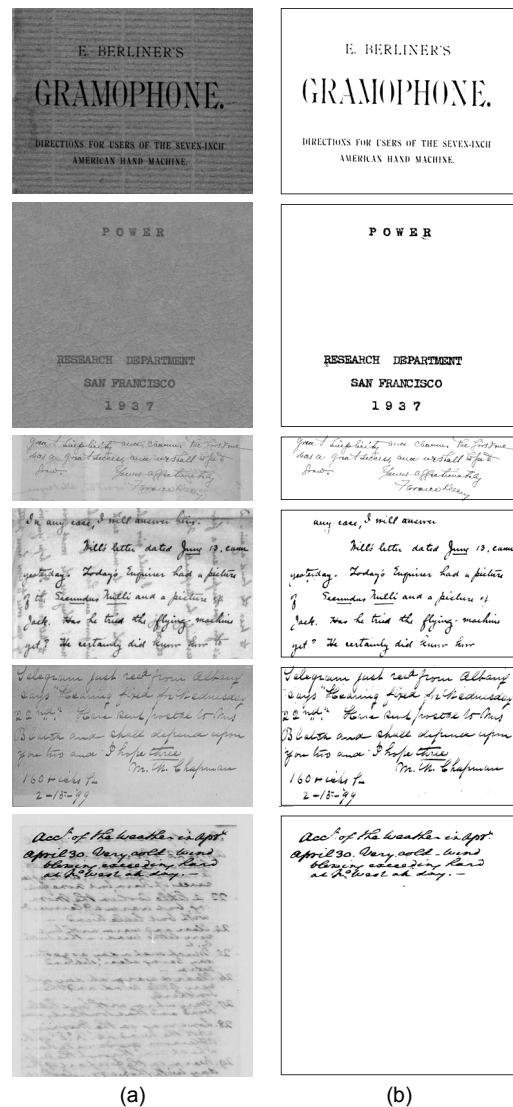


그림 7. 본 논문의 이진화 결과들 (a) 샘플 이미지 (b) 본 논문의 결과

Fig. 7 Binarization results of this paper (a) sample image (b) results of this paper

그림 7은 다양한 저하 이미지에 대한 본 논문의 결과를 보여준다. 그림 7을 통해 본 논문에서 제안한 방법이 여러 종류의 저하 이미지에서 좋은 이진화 결과를 보일 수 있음을 알 수 있다.

본 논문의 결과를 이전 연구들과 비교하기 위해 Su[7]등이 비교를 위해 사용한 Bickley diary dataset의 7개의 문서 이미지를 이용하고, 측정 방법으로는 DIBCO에서 사용하는 다음과 같이 정의되는 F-Measure를 사용하였다.

$$\begin{aligned} FM &= \frac{2 \times RC \times PR}{RC + PR} \\ RC &= \frac{CTP}{CTP + CFN} \\ PR &= \frac{CTP}{CTP + CFP} \end{aligned} \quad (6)$$

위 식에서 CTP, CFN, CFP는 각각 true positive, false negative, false positive에 해당하는 픽셀들의 수를 의미한다.

표 1. 이진화 방법들의 평가 결과

Table. 1 Evaluation results of binarization methods

이진화 방법	F-Measure(%)
Otsu[1]	50.42
Niblack[2]	67.71
Sauvola[3]	64.60
Gato[4]	69.13
Lu[5]	34.65
Su[6]	66.44
Su[7]	78.54
제안한 방법	69.12

표 1은 사용된 7개의 문서 이미지에 대한 F-Measure의 평균 값을 나타낸 것이다. 본 논문에서 제안한 방법이 Gato[4]등의 방법과 유사한 결과를 보이지만, Gato[4]등의 방법이 Sauvola[3]의 방법에 기반하기 때문에 Sauvola[3] 방법의 단점에서 자유롭지 못하다는 문제가 있다. Su[7]등의 방법이 실험에 사용된 문서 이미지에 대해 본 논문에서 제안한 방법보다 좋은 결과를 보여주지만, 이들의 방법은 다양한 텍스트 폭을 가지는 문서 이미지에 적합하지 않다는 단점을 가진다.

IV. 결 론

문서의 이진화는 문서 인식을 위한 이전 단계에서 사용되는 방법으로 문서 인식률과 연관되는 중요한 문제라 할 수 있다. 본 논문에서는 Matas등[8]이 제안한 MSER을 사용하여 문서 이미지를 이진화하는 방법을 제안하였다. 실험결과를 통해 본 논문에서 제안한 방법이 다양한 종류의 저하된 이미지에 적용될 수 있음을 알 수 있다.

본 논문에서 제안한 방법은 대비 이진 이미지가 실제 텍스트 영역에 포함되는 픽셀들 중 상당한 양을 놓치는 경우 좋지 않은 이진화 결과를 보일 수 있다. 따라서 실제 텍스트 픽셀들의 부분 집합을 얻는 방법으로 대비 이미지 외에 텍스트 경계 픽셀과 같이 다양한 방법을 이용함으로써 좀 더 안정적인 결과를 얻을 수 있을 것이다.

감사의 글

본 연구는 2014년도 부산외국어대학교 학술연구구조성비에 의해 이루어진 연구로서, 관계부처에 감사드립니다.

REFERENCES

- [1] N. Otsu, "A Threshold Selection Method from Gray Level Histogram," *IEEE Transactions on System, Man, Cybernetics*, vol. 19, no. 1, pp. 62-66, Jan. 1978.
- [2] W. Niblack, "An Introduction to Digital Image Processing," Englewood Cliffs, NJ:Prentice-Hall, 1986.
- [3] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225-236, Jan. 2000.
- [4] B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive Degraded Document Image Binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317-327, Mar. 2006.
- [5] S. Lu, B. Su, and C. L. Tan, "Document Image Binarization using Background Estimation and Stroke Edges," *International Journal on Document Analysis and Recognition*,

- vol. 13, no. 4, pp. 303-314, Dec. 2010.
- [6] B. Su, S. Lu, and C. L. Tan, "Binarization of Historical Handwritten Document Images using Local Maximum and Minimum Filter," in *Proceeding of the International Workshop on Document Analysis Systems*, pp. 159-166, Jun. 2010.
- [7] B. Su, S. Lu, and C. L. Tan, "Robust Document Image Binarization Technique for Degraded Document Images," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1408-1417, Apr. 2013.
- [8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," In *Proceeding of the British Machine Vision Conference*, pp. 384-393, 2002.



유영중(Young-Jung Yu)

부산외국어대학교 컴퓨터공학과 교수

※관심분야: 이미지 처리, 컴퓨터 그래픽스