한국산학기술학회논문지
Vol. 10, No. 2, pp. 358-363, 2009

에지 및 국부적 최소/최대 변환을 이용한 자연 이미지로부터 텍스트 영역 검출

박종천[1], 이근왕[2*]

Text Region Detection using Edge and Regional Minima/Maxima Transformation from Natural Scene Images

Jong-Cheon Park[1] and Keun-Wang Lee[2*]

요 약 자연이미지로부터 텍스트 영역 검출은 다양한 응용분야에 활용됨으로 이 분야의 많은 연구가 필요하다. 최근의 연구 방법은 에지 및 연결요소 기반 방법을 결합하는 다양한 알고리즘을 이용하여 텍스트 영역을 검출하고 있다. 그러므로 본 논문은 이러한 결합방법으로 에지 및 국부적 최소/최대 변환 방법을 이용하여 텍스트 영역을 검출하는 알고리즘을 제안한다. 명도 이미지로부터 에지 및 국부적 최소/최대 연결성분을 검출하고, 에지 및 국부적 최소/최대 연결성분을 레이블화한다. 레이블된 영역을 분석하여 텍스트 후보 영역을 검출하고, 검출된 각각의 텍스트 후보 영역을 결합하여 단일 텍스트 후보 이미지를 생성한다. 텍스트 후보 개별문자의 인접성 및 유사도를 비교하여 검증함으로서 최종적인 텍스트 영역을 검출한다. 실험결과 제안한 알고리즘은 에지 요소 및 국부적 최소/최대 연결요소 검출 방법을 결합하여 자연 이미지로부터 텍스트 영역 검출의 정확도 및 재현률을 향상할 수 있었다.

Abstract Text region detection from the natural scene images used in a variety of applications, many research are needed in this field. Recent research methods is to detect the text region using various algorithm which it is combination of edge based and connected component based. Therefore, this paper proposes an text region detection using edge and regional minima/maxima transformation algorithm from natural scene images, and then detect the connected components of edge and regional minima/maxima, labeling edge and regional minima/maxima connected components. Analysis the labeled regions and then detect a text candidate regions, each of detected text candidates combined and create a single text candidate image, Final text region validated by comparing the similarity and adjacency of individual characters, and then as the final text regions are detected. As the results of experiments, proposed algorithm improved the correctness of text regions detection using combined edge and regional minima/maxima connected components detection methods.

Key Words : Canny-Edge, Regional Minima/Maxima, Text Region Detection

1. 서 론

최근 모바일 기기 성능의 발전으로 이를 이용한 다양한 형태의 응용서비스가 연구되고 있으며, 그 중에서 모바일 기기에 내장된 카메라의 해상도가 높아짐에 따라 카메라로 획득된 이미지를 이용한 다양한 응용서비스의 필요성이 증가하고 있는 추세이다. 따라서 모바일 기기로 획득된 자연 이미지를 이용하여 기존의 고정되고 제안된 영역에서 획득된 이미지를 이용한 서비스로부터 다양한 환경에 노출된 자연 이미지로부터 필요한 정보를 추출하고자 하는 연구가 진행 중에 있다. 그 중에서 텍스트 정보는 데이터베이스, 교통정보, 관광정보, 등의 다양한 분

본 연구는 2008년도 청운대학교 학술연구조성비 지원에 의하여 연구되었음
[1]충북대학교 컴퓨공학과
[2]청운대학교 멀티미디어학과
*교신저자: 이근왕(kwlee@chungwoon.ac.kr)
접수일 09년 01월 05일 수정일 09년 02월 03일 게재확정일 09년 02월 18일

야에서 활용이 가능한 중요한 요소로 이용될 수 있다.

최근 연구 중에는 시각 장애인 안내를 위한 시각보조 시스템[1]이 개발 중에 있고, 외국인이 휴대장비(PDA, 스마트 폰)로 안내표지판, 교통표지판, 간판, 그리고 도로표지판 등의 각종 지리정보 및 관광정보를 텍스트 영역 검출과 언어 변환기술을 이용하여 사용자가 원하는 형태의 언어로 변환 가능하도록 하는 시스템[2]도 개발 중이다. 이 시스템은 휴대장비에 부착된 카메라로부터 획득한 자연 이미지에서 텍스트 정보를 추출하여 음성으로 텍스트 정보를 전달해주는 시스템이다. 이러한 시스템은 기본적으로 문자인식을 위한 전처리 과정으로 텍스트 영역을 정확히 추출하는 것을 요구한다. 그러나 자연 이미지에 포함된 텍스트는 복잡한 배경, 다양한 조명, 텍스트 크기, 텍스트의 방향, 그리고 다양한 컬러 분포 등의 특징을 갖고 있으므로 해결해야 할 많은 문제점을 갖고 있다.

2. 기존 연구 방법

텍스트 영역 검출 및 추출에 관한 많은 연구 논문이 발표되었으며, 텍스트 영역 검출 및 추출 방법은 영역-기반(region-based) 방법과 텍스쳐-기반(texture-based) 방법으로 분류되며 영역-기반방법은 연결요소 분석방법(Connected Component based method)과 에지-기반 분석방법(Edge-based method)으로 분류한다[3].

연결요소 분석 방법은 픽셀 값의 지역적 유사성을 바탕으로 그룹화 하는 과정을 수행함으로서 이미지에 존재하는 개별 영역을 식별하는 방법으로서 이와 관련된 아래에 제시된 기존의 연구 방법을 비교분석하였다.

장인영 등은 입력 이미지로 컬러 이미지를 이용하였고, 전처리는 컬러 이미지를 명도 이미지로 변환하고, 명암 대비 개선 및 적응적 임계값을 적용하였으며, 모폴로지 녹임(erosion)을 적용한 영상과 모폴로지(열림닫힘[OpenClose] + 닫힘열림[CloseOpen])/2가 적용된 이미지의 차이를 이용한 텍스트 영역을 추출방법이 연구되어졌다[4].

Y. Zhong 등은 컬러 이미지를 사용하였고, 전처리는 컬러 히스토그램을 계산하여 비슷한 컬러 영역별로 분할 합병(Split and Merge) 알고리즘을 적용하였다. 그리고 전처리 과정 중 삭제된 텍스트 영역을 복원한 후에 텍스트 영역을 추출하는 방법이 연구되어졌다[5].

L. Gu, T. Kaneko는 컬러 이미지를 이용하였고, 전처리는 컬러 이미지를 명도 이미지로 변환하고, DTT(Differential Top-hats)라는 모폴로지 분할 알고리즘과 방향성 필터(directional filter)를 사용하여 텍스트 영역과 배경 영역을 분리하는 방법을 적용하여 텍스트 영역을 추출하는 방법이 연구되어졌다[6].

에지-기반 방법과 관련된 연구를 살펴보면, Minhua Li, Chunheng Wang은 입력 이미지는 명도 이미지를 사용하였고, 배경 복잡도 분석과정을 수행하여 복잡도를 단순한 배경, 중간적 배경, 복잡한 배경 등으로 구분하여 각각의 복잡도에 맞는 파라미터로 에지를 검출하고 연결요소를 분석함으로서 텍스트 영역을 검출하는 방법이 연구되어졌다[7].

Toan Nguyen Dinh는 모바일 폰 카메라로 획득한 이미지를 입력이미지로 사용하였고, 입력 컬러 이미지를 명도 이미지로 변환하고, 수평방향 에지를 검출하고, 수평방향 히스토그램을 구하여 그 결과를 프로젝션함으로서 텍스트 영역을 검출하고, 검출된 텍스트 영역을 수직방향 히스토그램을 구하여 각각의 문자를 분리하였다[8].

에지-기반 방법은 이미지 대비의 높고 낮음에 따라 텍스트 영역에서 에지 검출의 정확성이 달라지고 복잡한 배경이 있는 경우 텍스트를 포함하지 않은 영역에도 에지가 검출되므로 텍스트 영역 검출 성능 및 검출속도를 저하시키는 결과를 초래한다. 반면 연결성분을 이용한 방법은 에지-기반 방법의 단점을 보완할 수 있다. 그러나 빛 또는 조명에 민감한 단점이 있으므로 이러한 문제점을 상호 보완하기 위한 방법으로 본 논문은 각각의 방법으로 검출한 결과를 결합함으로서 자연 이미지로부터 텍스트 영역을 검출하는 알고리즘을 제안하였다.

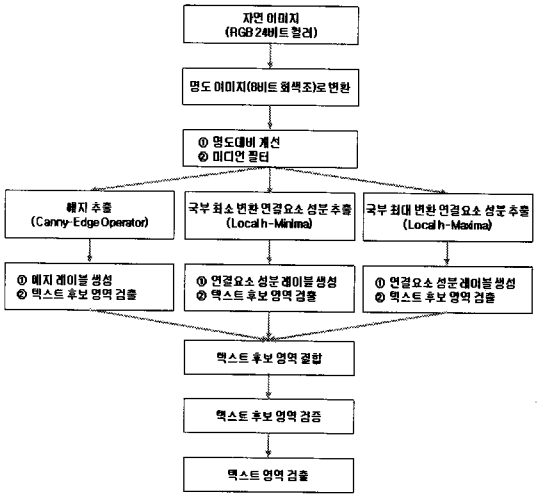
3. 제안 방법

텍스트 영역 검출은 그림 1과 같은 과정을 수행한다. 자연 이미지를 명도 이미지로 변환하고 전처리 과정을 수행하여 이미지를 개선한다. 텍스트 영역 추출은 에지특징과 연결요소 성분 특징을 이용하며 추출된 각각의 특징으로부터 텍스트 후보 영역을 추출하고 텍스트 후보 영역 검증을 수행하여 최종적인 텍스트 영역을 검출한다.

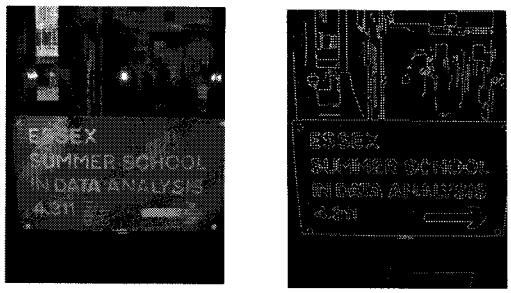
3.1 전처리 과정

디지털 카메라로 획득한 RGB 24비트 컬러 이미지를 명도 이미지(Gray-scale)로 변환하고, 전처리로서 명도대비를 개선하기위해 이미지의 명도 값에 따라 자동적으로 대비를 개선하는 절차를 수행한다. 이 과정으로 대비가 낮은 이미지에 포함된 텍스트 영역도 검출할 수 있도록 한다. 그리고 5×5 마스크의 크기를 이용하여 미디언 필터(Median)처리를 함으로서 이미지의 작은 잡음 성분을 제

거한다. 이 과정으로 연결요소 성분을 강화하여 연결요소 성분을 효과적으로 추출할 수 있도록 하였다.



[그림 1] 텍스트 영역 검출과정



(a) (b)

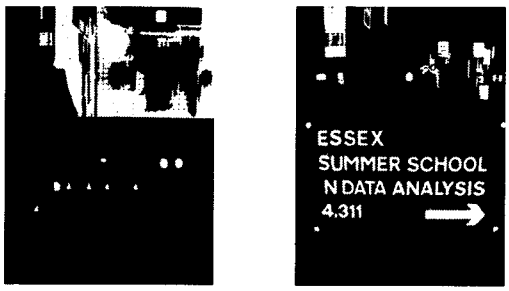
[그림 2] Canny 에지 검출기를 이용한 에지 성분 검출 결과 이미지

3.2 에지 성분 검출

명도 이미지로부터 에지 성분 검출은 Canny-Edge 검출기[9]를 이용하였고, 실험결과 에지 강도를 위한 임계값은 하한값은 0.1, 상한값은 0.2, 그리고 가우시안 필터 처리를 위한 시그마(sigma)값은 1.7로 설정하여 7×7 마스크 크기로 스무딩(Smoothing) 연산을 수행한다. 그림 2의 (a)는 명도 이미지, (b)는 Canny 에지 검출기를 이용하여 에지를 검출한 이미지를 보여준다.

3.3 연결요소 성분 검출

연결요소 성분 검출은 국부적 최소/최대 (Regional minima/maxima)변환을 수행하여 추출한다. 국부적 최소/최대 변환[10]은 명도 이미지를 대상으로 형태학적 재건(Morphological reconstruction)을 적용하여 연결요소 성분을 기반으로 반복하여 확장(dilation)연산을 수행함으로서 국부적으로 최소값과 최대값 영역을 검출한다. 국부적 최소/최대 변환을 위한 임계값은 실험결과 이미지 전체의 표준편차를 적용한 결과 텍스트 영역을 효과적으로 검출하였다. 그림 3은 국부적 최소/최대 변환을 적용하여 텍스트 영역의 연결요소 성분으로 추출한 결과이다. 그림 3의 (a)는 국부적 최소변환결과 이미지이고, (b) 국부적 최대변환 결과 이미지를 보여준다. 그림 3의 (b)의 국부적 최대 변환결과 대부분의 텍스트 영역을 검출하는 것을 알 수 있다. 텍스트 영역은 일반적으로 배경과 대비를 갖는 특징을 갖고 있으므로 국부적 최소/최대 변환으로 효과적으로 검출될 수 있다. 그러나 임계값에 따라서 일부 텍스트 영역은 검출하지 못하는 경우도 발생한다. 이러한 부분은 에지 검출 방법으로 해결될 수 있다.



(a) (b)

[그림 3] 국부적 최소/최대변환을 이용한 연결요소 성분으로 검출된 이미지

3.4 텍스트 후보 영역 검출

텍스트 후보 영역 검출은 에지 성분과 국부적 최소/최대 변환으로 검출된 연결요소를 바탕으로 텍스트 영역 특징을 만족하는 후보 개별문자를 추출하는 것이다. 이 과정을 위해서 에지 성분과 국부적 최소/최대 연결요소 성분을 레이블화하고 레이블된 영역의 특징을 검출하기 위해서 레이블된 영역의 구조적 특징을 이용한다. 레이블 영역의 구조적인 특징은 다음과 같은 조건을 만족하지 않아야 된다.

nr = 이미지의 높이, nc = 이미지의 너비
width = 레이블 영역의 너비
height = 레이블 영역의 높이
Number_Nested_Label
= 레이블 영역에 포함된 다른 레이블의 수

```
1) Max_Width_Height=max(width,height);
   if Max_Width_Height < 10
2) if width > height*3 || (width <= 4 &&
   height >= width*5) || height >= width*10
3) if Number_Nested_Label >= 10
4) if width < (nc*0.9) && height < (nr*0.9)
```

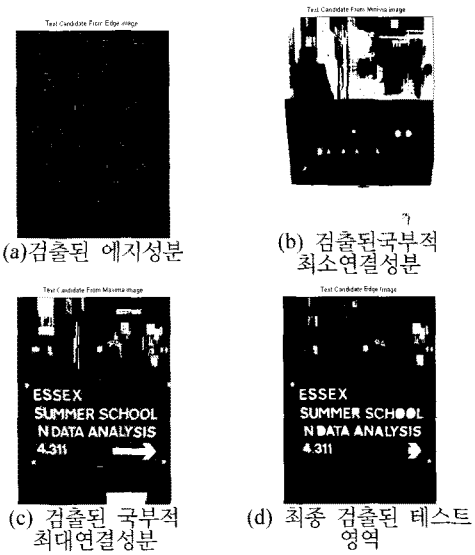
위의 조건 1)은 너비와 높이 중에서 가장 긴 것을 기준으로 최소 10픽셀 미만의 작은 텍스트 영역을 제외하는 것이고 조건 2)는 가로와 세로의 비율을 지정한 것이고, 조건 3)은 잡음 영역을 제외하는 것이고, 조건 4)는 너무 크거나 작은 텍스트 영역을 제외하는 것이다. 위와 같은 4가지 조건을 텍스트 영역의 후보를 검출하는 1차적인 조건이고, 추가적으로 에지 성분에 대한 조건은 에지 성분의 특징으로 아래와 같은 추가적인 조건을 만족하지 않는 것은 텍스트 후보 영역 개별문자에서 제외한다.

```
Total_Vertical_2_Lines
= 수직방향으로 검사한 Run의 길이가 2개 이상인 것
  의 합계
Total_Horizontal_2_Lines
= 수평방향으로 검사한 Run의 길이가 2개 이상인 것
  의 합계
1) Total_Vertical_2_Lines <= height * 0.5
2) Total_Horizontal_2_Lines <= width * 0.5
```

위의 조건은 텍스트 영역 에지 성분은 에지가 최소 2번 이상 존재해야만 한다는 조건을 반영한 것이다.

3.5 텍스트 후보 영역 결합 및 검증

에지 성분과 국부적 최소/최대 연결성분으로 검출된 텍스트 후보 개별문자 영역을 결합함으로서 하나의 텍스트 후보 영역 이미지를 생성하게 된다. 결합조건은 겹쳐진 영역이 2개 이상인 것을 조건으로 한다. 그림 4의 (d)는 (a)(b)(c)의 검출 결과를 결합하여 검출된 텍스트 후보 영역 이미지이다.

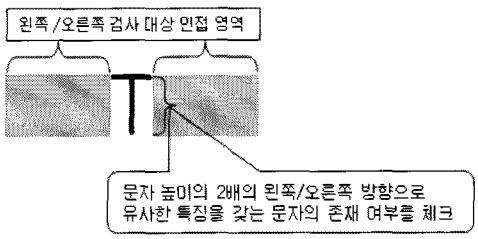


(a)검출된 에지성분 (b) 검출된국부적 최소연결성분

(c) 검출된 국부적 최대연결성분 (d) 최종 검출된 테스트 영역

[그림 4] 텍스트 후보 영역 추출 결과 이미지

텍스트 후보 영역 검증은 검출된 텍스트 후보 영역에서 개별 문자의 지역적 인접성을 검사함으로서 해당하는 개별문자의 텍스트 영역 여부를 결정한다. 그림 5는 텍스트 후보 영역 검증을 위한 개별 문자 인접성을 체크하는 것을 보여준다. 인접된 문자의 유사성의 기준은 아래와 같다.

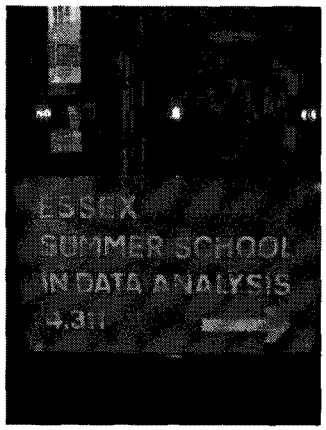


[그림 5] 텍스트 영역 검증을 위한 인접성 체크

```
Width = 검사대상 문자의 너비
Height = 검사대상 문자의 높이
NB_Width = 인접한 텍스트 후보 영역의 너비
NB_Height = 인접한 텍스트 후보 영역의 높이
Similarity_Width = NB_Width/width;
Similarity_Height = NB_Height/height;
NB_Similarity =
  max(Similarity_Width, Similarity_Width);
if height >= width*5 &&
  Similarity_Height >= 0.7 &&
```

```
    Similarity_Height <= 1.5
elseif Similarity_Height >= 0.7 &&
       Similarity_Height <= 1.5 &&
       Similarity_Width <= 2
```

그림 6은 텍스트 영역 검증 결과 검출된 텍스트 영역 이미지를 보여준다.



[그림 6] 텍스트 영역 검출 결과 이미지

4. 실험 결과

실험 이미지는 ICDAR 2003 표준 이미지이고, 텍스트 영역 검출률 성능평가는 재현률(Recall)과 정확률(Precision)을 평가요소[11]로 하였고, 표 1과 같이 제시하였다.

그림 6의 실험결과를 보면 상단의 텍스트 영역이 아닌 부분을 텍스트 영역으로 검출하는 것을 볼 수 있는데 이것은 텍스트 영역을 결정하는 구조적 특징이 제안한 방법의 조건을 만족함으로서 검출되는 것으로 2차적인 텍스트 영역의 구조적 특징 요소가 필요하다. 그리고 그림 6의 문자 중에서 영문자 "I"는 검출하지 못한 것은 이미지의 대비가 상대적으로 낮은 영역의 국부적 최소/최대 연결성분 검출과정에서 임계값의 영향으로 검출하지 못한 결과이다.

Sum : 실험 대상 이미지 내에 존재하는 전체 텍스트 영역의 문자 수

True : 정확히 검출한 텍스트 영역의 문자수

Error : 검출하지 못한 텍스트 영역의 문자 수

False : 텍스트 영역이 아닌 영역을 텍스트 영역으로 잘못 검출한 문자 수

Precision : True / (True+False+Error)

Recall : True / Sum

[표 1] 텍스트 영역 검출률

실험 이미지	Sum	True	Error	False	Precision	Recall
안내판	1070	779	35	74	87.7	72.8
간판	1213	853	39	55	90.1	70.3
교통 표지판	512	345	15	50	84.1	67.4
번호판	545	321	30	50	80.0	58.9
교통 안내판	453	331	23	50	81.9	73.1
광고판	232	220	19	11	88.0	94.8
전체	4025	2849	161	290	86.3	70.8

실험결과 배경이 비교적 단순한 간판, 광고판, 안내 표지판은 다른 이미지에 비해 상대적으로 높은 정확도를 보였고, 배경이 다양하고 복잡한 교통 표지판과 교통 안내판은 낮은 Precision과 Recall 값을 나타냈다.

5. 결론 및 향후 연구방향

본 논문은 자연 이미지로부터 에지 및 연결요소 성분을 결합하여 텍스트 영역을 검출하는 방법을 제안함으로서 기존 연구의 문제점을 해결 할 수 있었다. 에지 성분으로 검출되지 않는 부분은 국부적 최소/최대 변환 연결성분을 검출함으로 텍스트 영역을 검출 하였고, 국부적 최소/최대 변환을 위한 임계값의 영향으로 검출되지 않는 텍스트 영역은 에지 성분으로 검출함으로서 텍스트 영역을 효과적으로 검출 할 수 있었다.

향후 연구 과제는 텍스트 영역 검출 성능 평가 요소 중 Error 최소화하기 위한 방법으로 에지 성분 검출을 위한 파라미터 설정값을 결정하는 방법과 국부적 최소/최대 변환을 위한 최적의 임계값을 설정하는 연구가 필요하다. 그리고 False 요소를 최소화하기 위해서 텍스트 영역의 군집특징과 텍스트 영역의 특징을 결정하는 구조적인 특징 요소를 도출하는 방법을 연구하는 것이다. 마지막으로 색 정보를 추가하여 효과적으로 결합하는 방법에 관한 연구가 필요하다.

참 고 문 헌

[1] N. Ezaki, M. Bulacu, L. Schomaker, "Text detection from natural scene images: towards a system for

visually impaired persons", Pattern Recognition, ICPR 2004, Proceedings of the 17th International Conference on Volume 2, pp.683-686, 2004.

[2] J. Yang, X. Chen, J. Zhang, Y. Zhang, A. Waibel, "Automatic detection and translation of text from natural scenes", Acoustics, Speech, and Signal Processing, IEEE International Conference on Volume 2, pp.2101-2104, 2002.

[3] K. Jung, K. I. Kim and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", Pattern Recognition, Volume 37, Issue 5, May pp.977-997, 2004.

[4] 장인영, 고병철, 변혜란, "모폴로지를 이용한 비디오 영상에서의 자동 문자 추출," 정보과학회 추계학술대회 논문집, pp.418-420, 2001.

[5] Y. Zhong, K. Karu, A. K. Jain, "Locating Text in Complex Color Images", Pattern Recognition, Vol. 28. No. 10, pp. 1523-1535, 1995.

[6] L. Gu, T. Kaneko, "Robust extraction of Characters from color scene Images Using Mathematical morphology", Proc. 7th Int. Conf. Pattern Recognition, Vol. 2, pp.1002-1004, 1998.

[7] M.A. Smith and T. Kanade, "Video Skimming for Quick Browsing Based on Audio and Image Characterization", Carnegie Mellon University, Technical Report CMU-CS-95-186, 1995.

[8] D. Chen, K. Shearer, and H. Bourlard, "Text Enhancement with Asymmetric Filter for Video OCR", Proc. of International Conference on Image Analysis and Processing, pp.192-197, 2001.

[9] W. Jiang , Q. Shao-Lin , Q. Zhuo, W. WenYuan , "Automatic text detection in complex color image", Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on Volume 3, pp.1167-1171, 2002.

[10] H. Xian-Sheng, Z. Liu Wenyin, Hong-Jiang, "An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms", IEEE Transactions on Circuits and Systems for Video Technology, Vol 14, No. 4, pp.498-507, 2004.

[11] Minhua Li, Chunheng Wang, "An Adaptive Text Detection Approach in Images and Video Frames", Neural Networks, IJCNN 2008, pp.72-77, 2008.

[12] Toan Nguyen Dinh, Jonghyun Park, GueeSang Lee, "Low-Complexity Text Extraction in Korean Signboard for Mobile Applications", CIT(Computer and Information Technology), 2008. 8th IEEE International Conference on pp.333-337, 2008

[13] J. Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 6, pp.679-698, 1986.

[14] Luc Vincent, "Morphological Gray-scale Reconstruction in Image Analysis:Applications and Efficient Algorithms", Image Processing, IEEE Transactions on Volume 2, pp.176-201, 1993.

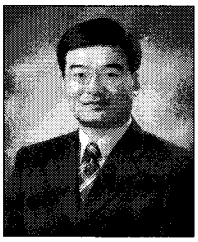
[15] V. Raghavan, P. Bollmann, and G. Jung, "A critical investigation of recall and precision as measures of retrieval system performance," ACM Trans. on Information Systems, Vol. 7, pp.205-229, 1989.

박 종 천(Jong-Cheon Park) **[정회원]**

- 1994년 2월 : 한밭대학교 전자계산학과(공학사)
- 1998년 2월 : 숭실대학교 전자계산학과 (공학석사)
- 2004년 8월 : 충북대학교 컴퓨터공학과 박사과정 수료
- 2007년 9월 ~ 현재 : 충북대학 시간강사

<관심분야>
컴퓨터 비전, 영상처리, 인공지능

이 근 왕(Keun-Wang Lee) **[종신회원]**



- 1993년 2월 : 한밭대학교 전자계산학과(공학사)
- 1996년 2월 : 숭실대학교 컴퓨터학과(공학석사)
- 2000년 2월 : 숭실대학교 컴퓨터학과(공학박사)
- 2001년 ~ 현재 : 청운대학교 멀티미디어학과 부교수

<관심분야>
멀티미디어 통신, 멀티미디어 응용, 교육 콘텐츠