

에지 및 컬러 양자화를 이용한 모바일 폰 카메라 기반 장면 텍스트 검출

박종천¹, 이근왕^{2*}

¹충북대학교 컴퓨터공학과, ²청운대학교 멀티미디어학과

Mobile Phone Camera Based Scene Text Detection Using Edge and Color Quantization

Jong-Cheon Park¹ and Keun-Wang Lee^{2*}

¹Division of Computer Engineering, Chungbuk National University

²Dept. of the Multimedia Science, Chungwoon University

요약 자연 영상 내에 포함된 텍스트는 영상의 다양하고 중요한 특징을 갖는다. 그러므로 텍스트를 검출하고 추출하여 인식하는 것이 중요한 연구대상으로 연구되고 있다. 최근 모바일 폰 카메라를 기반으로 다양한 분야에서 많은 응용 기술이 연구 개발되고 있다. 본 논문은 에지 및 연결요소를 이용한 장면 텍스트 검출 방법을 제안한다. 그레이 스케일 영상으로부터 에지 성분 검출과 지역적 표준편차를 이용하여 텍스트 영역의 경계선을 검출하고, RGB 컬러 공간의 유클리디안 거리를 기준으로 연결요소를 검출한다. 검출된 에지 및 연결요소를 레이블링하고 각각 영역의 외곽사각형을 구한다. 텍스트의 휴리스틱 이용하여 후보 텍스트를 추출한다. 후보 텍스트 영역을 병합하여 하나의 후보 텍스트 영역을 생성하고, 후보 텍스트의 지역적 인접성과 구조적 유사성으로 후보 텍스트를 검증함으로써 최종적인 텍스트 영역을 검출하였다. 실험결과 에지 및 컬러 연결요소 특징을 상호 보완함으로써 텍스트 영역의 검출률을 향상시켰다.

Abstract Text in natural images has a various and important feature of image. Therefore, to detect text and extraction of text, recognizing it is a studied as an important research area. Lately, many applications of various fields is being developed based on mobile phone camera technology. Detecting edge component from gray-scale image and detect an boundary of text regions by local standard deviation and get an connected components using Euclidean distance of RGB color space. Labeling the detected edges and connected component and get bounding boxes each regions. Candidate of text achieved with heuristic rule of text. Detected candidate text regions was merged for generation for one candidate text region, then text region detected with verifying candidate text region using ectilarity characterization of adjacency and ectilarity between candidate text regions. Experctental results, We improved text region detection rate using complementary of edge and color connected component.

Key Words : Canny-Edge, Euclidean Distance, Standard Deviation, Text Detection

1. 서론

최근 디지털 카메라, 모바일 카메라, PDA, 스마트 폰 등 개인 정보 단말기의 사용이 보편화 되면서 이와 관련

된 응용 기술이 급속한 성장을 하고 있으며, 앞으로 더 많은 응용 기술과 부가가치를 창출할 수 있을 것으로 전망하고 있다. 그 중에서 특히 스마트 폰을 이용한 응용기술이 핫 이슈로서, 전 세계적으로 스마트 폰 시장의 경쟁

본 연구는 2009년도 청운대학교 학술연구조성비 지원에 의하여 연구되었음.

*교신저자 : 이근왕(kwlee@chungwoon.ac.kr)

접수일 10년 02월 19일

수정일 10년 03월 17일

게재확정일 10년 03월 18일

이 시작되어 다양한 제품이 출시되고 있으며, 그에 따른 다양한 소프트웨어 개발이 이루어지고 있다. 스마트 폰에 장착된 카메라를 이용한 다양한 응용프로그램이 개발되고 있으며, 그 중에서 상표 및 명함의 글자 등을 인식하여 그 결과를 인터넷 검색으로 연결함으로써 인터넷 검색 환경을 개선하여 사용자의 인터넷 검색을 기존의 텍스트 기반에서 멀티미디어 기반으로 확장하는 서비스를 제공하고 있다. 모바일 카메라 기반 연구는 자동 주행 시스템, 시각 장애인 보행 지원 시스템[1], 지능형 운전 시스템, 박물관 정보 안내 시스템, 도서 검색 시스템, 거리 간판 인식을 위한 내비게이션 시스템 등과 같은 응용분야에서 연구 중에 있다.

텍스트 영역 검출 방법의 기존의 연구는 대부분 배경과 텍스트 영역이 차이가 뚜렷한 영상에서는 정확한 텍스트 영역을 검출 하였으나 배경이 복잡하고 대비가 낮은 영상에 포함된 텍스트는 검출률이 떨어지는 결과를 보였다. 자연 영상에 포함된 텍스트는 다양한 조명 변화와 그에 따른 그림자, 텍스트의 크기와 종류, 방향, 다양한 컬러 분포 등의 특징으로 인하여 텍스트 영역을 정확히 검출하기 어려운 문제점이 있다.

문자영역 검출 방법은 영역-기반 방법(Region-based method)과 질감-기반 방법(Texture-based method) 으로 크게 분류할 수 있는데 영역-기반 방법은 텍스트 영역의 컬러 정보, 밝기 정보, 텍스트와 배경 영역이 갖는 특징의 차이를 이용하여 텍스트 영역과 배경 영역을 분리함으로써 텍스트 영역을 검출하는 방법으로 이는 다시 에지-기반 방법과 연결요소-기반 방법으로 분류된다. 에지-기반 방법은 텍스트 영역의 에지를 검출하고 그 특징을 분석함으로써 텍스트 영역을 검출하는 방법으로, 이와 관련된 연구를 살펴보면,

Toan Nguyen Dinh는 모바일 폰 카메라로 획득한 컬러 영상을 그레이스케일 영상으로 변환하고, 수평방향 에지를 검출 및 히스토그램을 구하여 그 결과를 투영함으로써 텍스트 영역을 검출하였으며, 검출된 텍스트 영역을 다시 수직방향 히스토그램을 구하여 각각의 텍스트 영역의 문자를 분리하였다[2].

Minhua와 Chunheng은 그레이스케일 영상의 배경 복잡도 분석과정을 수행하여 배경 복잡도가 단순 배경, 중간복잡도 배경, 복잡한 배경 등으로 구분하여 각각의 복잡도에 따른 파라미터로 에지를 검출하고 연결요소를 분석함으로써 텍스트 영역을 검출하였다[3].

Smith와 Kanade는 입력 영상을 3×3 수평성분 검출 필터를 적용하고 임계값을 사용하여 수직에지 성분을 검출하고 스무딩 연산을 수행하여 작은 에지 성분은 제거하고, 끊어진 에지 성분은 연결하여 텍스트 영역을 검출 하

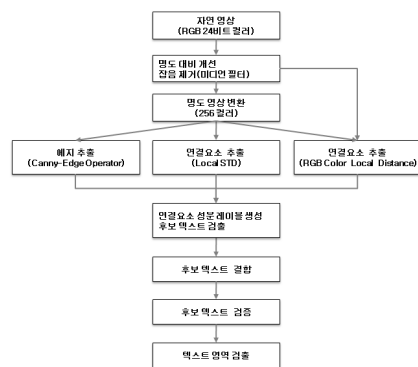
였다[4].

연결요소-기반 방법은 픽셀값의 지역적 유사성을 바탕으로 그룹화 하는 과정을 수행함으로써 영상 전체에 존재하는 개별 영역을 식별하는 방법으로 기존의 관련된 연구들을 보면, Jain과 Yu은 입력영상으로 정지 영상, 웹 영상, 컬러 영상 및 동영상에서 문자를 검출하는 방법을 제안하였다[5]. 정지 영상과 웹 영상에서는 문자의 밝기 값이 균일하다는 특징을 이용하고, 컬러 영상 및 동영상에 대해서는 컬러 연속성을 이용하고, 전자의 전처리 과정으로 밝기 값이 균일한 특징을 이용하여 다중치 분해(multi-valued decomposition)로 영역을 검출하고, 후자의 컬러 연속성을 이용하여 색조임을 수행한 후 다중치 분해로 문자와 배경을 분리한 연결요소를 분석하여 텍스트 영역을 검출하였다.

Wolf 등은 입력 영상으로부터 경계선을 검출하고, 가로방향의 저주파 필터링 및 형태학적 연산을 이용하여 연결 성분을 얻고, SVM(Support Vector Machine)을 이용하여 텍스트와 배경을 구분하여 텍스트 영역을 검출 하였다[6].

기존의 텍스트 영역 검출에 관한 연구 방법은 대부분 제한된 환경에서 얻어진 영상의 텍스트 영역 검출률은 높았으나, 다양한 외부 환경과 복잡한 배경을 갖는 영상에 포함된 텍스트의 경우는 검출률이 낮았다.

에지-기반 방법은 영상 대비에 따라 텍스트 영역 에지 검출률의 정확성이 달라지고, 복잡한 배경이 있는 경우 텍스트 영역이 아닌 영역에서 텍스트를 검출함으로써 검출률 및 검출속도를 저하시키는 결과를 초래한다. 반면, 연결요소-기반 방법은 에지-기반 방법의 문제점을 보완할 수 있으나 빛과 조명에 민감한 문제점이 있다. 따라서 이러한 문제점을 상호 보완하기 위한 방법으로 본 논문은 각각의 방법을 조합하고 결합함으로써 자연 영상에서 텍스트 영역을 검출하는 알고리즘을 제안한다.



[그림 1] 제안한 텍스트 영역 검출도

2. 제안 방법

그림 1에 제시한 과정을 수행함으로써 텍스트 영역을 검출한다. 모바일 폰 카메라로 획득된 24비트 컬러를 갖는 자연 영상을 그레이스케일 영상으로 변환하고 전처리 과정을 수행하여 영상 개선 및 잡음을 제거한다. 그레이스케일 영상으로부터 에지를 검출하고, 지역적 표준편차를 이용하여 텍스트 영역 경계의 연결성분을 검출한다. 컬러 영상으로부터 RGB 유클리디안 거리를 기준으로 연결요소를 검출한다. 각각의 결과를 레이블링하여 외각사각형을 구하고 이를 분석하여 후보 텍스트를 선정한다. 선정된 후보 텍스트를 조합 및 결합하여 텍스트 영역을 설정하고 텍스트의 지역성과 구조적 특징을 휴리스틱을 이용하여 검증함으로써 최종적인 텍스트 영역을 검출한다.

2.1 전처리

RGB 24비트 컬러 영상의 명도대비 개선을 수행함으로써 대비가 낮은 영상의 대비를 강화한다. 그리고 5×5 마스크 크기를 갖는 미디언 필터(Median)를 적용함으로써 영상의 잡음 크기를 제거한다. 그리고 그레이스케일 영상으로 변환하여 에지 성분을 검출할 수 있도록 한다.

2.2 에지 성분 검출

자연영상에서 텍스트는 높은 대비를 갖는 특징을 지닌다. 이러한 특징으로 텍스트와 배경의 밝기 차이가 발생함으로써 텍스트 외곽에 에지성분이 존재하게 된다. 그러므로 에지 성분을 검출함으로써 텍스트의 특징을 검출한다. 에지 성분 검출은 Canny-Edge 검출기[7]를 이용하였고, 검출을 위한 파라미터는 실험결과 임계값은 0.3, 시그마(sigma)값은 1.0으로 설정하였다. 그림 2는 에지 검출 결과를 보여준다. 이 경우 텍스트 영역이 빛을 등지고 있으므로 대비가 낮은 상태로 존재함으로써 에지 성분을 검출하지 못하는 것을 알 수 있다.



(a)원 영상 (b)Canny-Edge 검출기
[그림 2] 에지 성분 검출 영상

2.3 연결요소 검출

에지 검출결과 대비가 낮은 경우 텍스트 영역의 에지

성분을 검출하지 못하는 문제점을 보완하기 위해서 지역적 연결성을 고려한 연결요소 성분을 검출한다. 연결요소 성분 검출은 컬러 영상과 그레이스케일 영상 각각에 대해서 수행된다. 먼저 그레이스케일 영상에 대해서는 지역적 표준편차를 구하고 이를 임계값을 기준으로 연결요소 성분을 얻는다. 지역적인 3×3 영상의 중심 픽셀을 기준으로 이웃한 8 방향 픽셀에 대한 표준편차를 식 1로 구한다. 텍스트와 배경의 밝기 차이로 인해서 표준편차가 텍스트 외곽선에서 상대적으로 텍스트 내부보다 큰 값을 갖게 된다.

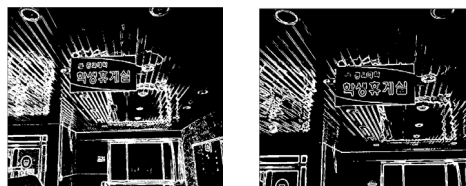
$$s = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2 \quad \text{단,} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

산출된 표준편차를 영상 전체의 표준편차에 따라 임계값을 설정하여 연결요소 성분을 검출한다. 실험결과 임계값은 영상전체의 표준편차가 70보다 크면 5로 설정하고 작으면 15로 설정하였다.

RGB 컬러 영상으로부터 RGB 컬러의 유클리디안 거리를 산출함으로써 연결요소 성분을 검출한다. 식2는 RGB 컬러의 유클리디안 거리를 계산하는 수식을 보여준다.

$$\|c_1 - c_2\| = \sqrt{(c_{1,x} - c_{2,x})^2 + (c_{1,y} - c_{2,y})^2 + (c_{1,z} - c_{2,z})^2} \quad (2)$$

RGB 컬러의 유클리디안 거리는 인접한 8 방향에 대해서 모두 구하여 평균거리를 산출한다. 임계값은 표준편차와 같은 값을 이용하여 연결요소 성분을 검출하였다. 산출된 평균 거리를 임계값을 기준으로 텍스트 영역과 비 텍스트 영역을 구분하여 이진 영상을 생성한다. 그림 3은 지역적 표준편차 및 RGB 컬러 유클리디안 거리를 이용한 연결요소 검출 결과 영상을 보여준다. 에지 검출 수행결과 검출하지 못한 텍스트를 검출하는 것을 알 수 있다.



(a)표준편차 (b)RGB 거리
[그림 3] 연결요소 성분 검출 영상

2.4 후보 텍스트 검출

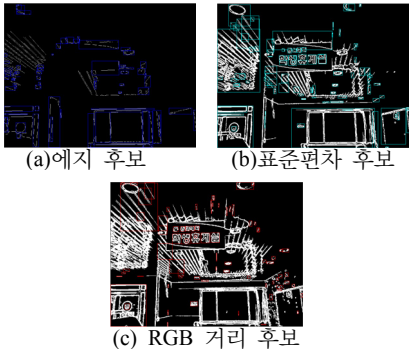
검출된 에지 및 연결요소 성분을 레이블링하고 외곽사각형 영역을 구한다[8]. 텍스트의 특징을 휴리스틱을 적

용하고 이를 분석하여 후보 텍스트를 선정하였다. 레이블링 과정은 에지 및 연결요소 검출과정에서 이진 영상으로부터 8 방향으로 연결된 성분을 모두 찾아 번호를 할당하는 작업을 수행하게 된다. 레이블 영역의 크기가 30 픽셀 미만이면 텍스트를 구성할 수 없으므로 제거하였다. 레이블 영역의 구조적 특징은 표 2에 제시하였다. 표 2에 제시한 조건은 텍스트 영역 에지는 텍스트 에지 구조의 특징으로 최소 2번 이상 존재해야만 한다는 것을 반영한 것이다.

[표 1] 텍스트 레이블 영역의 구조적 조건

- 1) Total_Vertical_2_Lines <= height * 0.5
 - 2) Total_Horizontal_2_Lines <= width * 0.5
- 단, Total_Vertical_2_Lines는 수직방향으로 검사한 Run의 길이가 2개 이상인 것의 수
Total_Horizontal_2_Lines는 수평방향으로 검사한 Run의 길이가 2개 이상인 것의 수

그림 4는 에지 성분, 지역적 표준편차 연결요소 성분 그리고 RGB 컬러 유클리디안 거리를 이용하여 검출된 후보텍스트 영상을 각각 보여준다. 그림 4의 (a)~(c) 모두 텍스트 영역이 아닌 영역에서도 텍스트를 검출하는 것을 볼 수 있다. 따라서 텍스트 영역의 구조적 특징을 적용하여 이러한 영역을 제거한다. 이를 위한 조건은 표 3에 제시하였다.



[그림 4] 후보 텍스트 검출

[표 2] 텍스트 영역 구조적 조건

- nr는 영상의 높이, nc는 영상의 너비
width는 레이블 영역의 너비
height는 레이블 영역의 높이
Number_Nested_Label은 레이블 영역에 포함된 다른 레이블의 수
- 1) Max_Width_Height=max(width,height);
Max_Width_Height < 10
 - 2) width > height*3 or (width <= 4 and height >= width*5) or height >= width*10
 - 3) Number_Nested_Label >= 10
 - 4) width < (nc*0.9) and 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 텍스트의 문자가 갖는 가로와 세로의 비율이고, 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, 2차 조건으로 표 4에 제시한 지역성과 구조적 유사도를 적용하여 조건을 만족하지 않는 것은 후보 텍스트 개별문자에서 제외한다.

[표 3] 텍스트 지역성과 구조적 유사도

- NB_Width = 인접한 후보 텍스트 영역의 너비
NB_Height = 인접한 후보 텍스트 영역의 높이
Similarity_Width = NB_Width/width;
Similarity_Height = NB_Height/height;
NB_Similarity =
max(Similarity_Width, Similarity_Width);
If height >= width*5 and
Similarity_Height >= 0.7 and
Similarity_Height <= 1.5
else if Similarity_Height >= 0.7 and
Similarity_Height <= 1.5 and
Similarity_Width <= 2

2.5 후보 텍스트 결합 및 검증

검출한 에지 성분과 지역적 표준편차 그리고 RGB 거리를 이용한 연결요소 성분으로 검출된 후보 텍스트 영역에 가중치를 부여하여 서로 합병함으로써 후보 텍스트 영역을 선정한다. 선정된 후보 텍스트 영역의 에지의 모양과 텍스트 영역 수평/수직 프로파일 특징을 이용하여 후보 텍스트 영역을 검증한다. 합병 조건은 세 가지 방법으로 검출된 영역이 2 이상일 때 겹쳐지는 영역으로 하였다.

후보 텍스트 검증은 표 4에서 제시한 지역성 및 구조적 유사성을 기준으로 하였다. 그림 5는 텍스트 영역 검증 결과 검출된 텍스트 영역 영상을 보여준다.

3. 실험 결과

본 논문에서는 텍스트 영역 검출물의 성능을 평가하기 위해 다양한 환경에서 모바일 폰 카메라를 이용하여 촬영된 영상을 사용하였고, 평가방법은 아래와 같은 식을 이용하여 정확률(Precision)과 재현율(Recall)로 평가하였다[9].

$$\text{Precision} = \frac{\text{True}}{\text{True} + \text{Part} + \text{Error}} \quad (3)$$

$$\text{Recall} = \frac{\text{True}}{\text{Sum}} \quad (4)$$

$$\text{Sum} = (\text{True} + \text{Part} + \text{Missing}) \quad (5)$$

True는 정확히 검출한 텍스트 영역의 수, Part는 텍스트의 일부분만을 검출한 수, Error는 텍스트가 아닌 영역을 텍스트로 검출한 수, Missing은 텍스트를 검출하지 못한 수를 의미한다.

그림 5의 실험결과를 보면 왼쪽 상단 및 오른쪽 하단에서 텍스트 영역이 아닌 부분을 텍스트 영역으로 검출하는 것을 볼 수 있다. 이것은 텍스트 영역을 결정하는 구조적 특징이 제안한 방법의 조건을 만족함으로써 검출되는 것으로 2차적인 텍스트 영역의 구조적 특징을 적용하여 보완할 필요성이 있다.

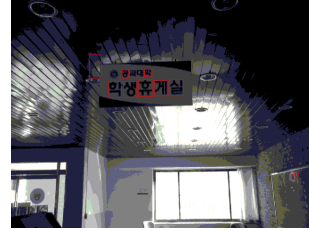
[표 4] 텍스트 영역 검출률

영상종류	Sum	True	Error	Part	Missing
안내표지	1070	808	35	100	162
간판	1213	930	66	65	218
교통표지	512	375	25	63	74
번호판	545	450	40	60	35
교통안내	453	370	10	33	50
광고	232	209	19	18	5

영상종류	Precision	Recalling
안내표지	85.7	75.5
간판	87.7	76.7
교통표지	81.0	73.2
번호판	81.8	82.6
교통안내	89.6	81.7
광고	85.0	90.1

실험결과 배경이 비교적 단순한 간판, 광고판, 안내 표지판은 다른 영상에 비해 상대적으로 높은 정확도를 보였고, 배경이 다양하고 복잡한 교통 표지판과 번호판은

낮은 Precision과 Recall 값을 나타냈다.



[그림 5] 텍스트 영역 검출 결과 영상

5. 결론 및 향후 연구방향

본 논문에서는 모바일 폰 카메라로부터 획득한 자연 영상에 존재하는 텍스트 영역을 효율적으로 검출하기 위해 에지 성분과 지역적 표준편차 그리고 RGB 컬러 유클리디안 거리를 이용한 연결요소 성분의 특징으로 텍스트 영역을 검출하는 방법을 제안하였다. 각각의 방법으로 검출된 영역의 조합하여 최종적인 텍스트 영역을 검출하였다. 따라서 각각의 방법의 문제점을 보완할 수 있었다.

텍스트가 아닌 영역을 제거를 위한 임계값을 결정하는 파라미터는 영상을 얻는 시점의 날씨, 조명, 그리고 시간 등에 따라 적절하게 설정되어야 함으로 이에 대한 데이터를 충분히 얻어야 하고 적절한 파라미터의 설정에 관한 연구가 필요하다. 추가적으로 모바일 폰에 내장될 것을 고려하여 알고리즘을 최적화하는 연구도 요구된다.

참고문헌

- [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] 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
- [3] Minhua Li, Chunheng Wang, "An Adaptive Text Detection Approach in Images and Video Frames", Neural Networks, IJCNN 2008, pp.72-77, 2008.
- [4] Smith, M. A. and T. Kanade, "Video Skimming for Quick Browsing Based on Audio and Image

Characterization”, Carnegie Mellon University, Technical Report CMU-CS-95-186, 1995.

- [5] Anil K. Jain, Bin Yu, “Automatic Text Location in Images and Video Frames”, Pattern Recognition, Vol. 31, No. 12, pp. 2055-2076, 1998.
- [6] Wolf, C., J.M. Jolion and F. Chassaing, “Text Localization, Enhancement and Binarization in Multimedia Documents,” In Proceedings of the International Conference on Pattern Recognition, Vol. 4, pp. 1037-1040, 2002.
- [7] 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.
- [8] Haralick, Robert M., and Linda G. Shapiro, “Computer and Robot Vision”, Vol 1, Addison-Wesley, pp.28-48, 1992.
- [9] Junker, M. and R. Hoch, “On the Evaluation of Document analysis components by recall, precision, and accuracy”, Proceedings of ICDAR, pp.713-716, 1999.

이 근 왕(Keun-Wang Lee)

[정회원]



- 1993년 2월 : 한밭대학교 전자계산학과 (공학사)
- 1996년 2월 : 숭실대학교 컴퓨터학과 (공학석사)
- 2000년 8월 : 숭실대학교 컴퓨터공학과 (공학박사)
- 2001년 2월 ~ 현재 : 청운대학교 멀티미디어학과 부교수

<관심분야>

멀티미디어 통신, 멀티미디어 응용, 교육 콘텐츠

박 종 천(Jong-Cheon Park)

[정회원]



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

<관심분야>

컴퓨터비전, 영상처리, 인공지능