

모바일 환경에서 실시간 영상 전경 추출 연구

윤종철*

Real-time Video Matting for Mobile Device

Jong-Chul Yoon*

요약 최근 모바일 기기의 영상 촬영 기능의 확대에 따라, 영상처리를 위한 다양한 어플리케이션들이 모바일 환경으로 이식되고 있다. 또한 모바일 기기의 컴퓨팅 능력또한 상승으로 기존에 컴퓨터에서만 가능하였던 다양한 영상처리 기술들이 모바일 환경으로 이식되고 있다. 이러한 영상처리 기술 중, 사용자가 촬영한 영상에서 전경만을 추출하여 원하는 새로운 배경에 합성하는 문제는 다양한 어플리케이션에서 사용이 가능하나 계산이 복잡하다는 문제점으로 모바일 환경 이식에 어려움을 가지고 있었다. 본 연구에서는 모바일 기기로 촬영된 영상을 전경/배경으로 구분하여 목표 영상에 실시간으로 합성할 수 있는 영상합성 기술을 제안한다. 사용자가 촬영한 영상에서 배경이 움직이지 않는다는 가정하에 자동으로 전경을 추출하며 이를 새로운 배경에 합성하는 기법을 소개한다. 모바일 촬영의 특성을 고려하여 촬영시 약한 움직임을 포함하는 영상에서의 자동 전경 추출 알고리즘을 개발하며 이를 SIMD 및 GPGPU기반의 가속화 알고리즘을 사용하여 SD급 화질의 영상에 대해 모바일 상에서 실시간 처리가 가능한 결과를 보인다. 본 논문의 기술을 사용하여 상용화 가능한 영상처리 어플리케이션의 개발이 가능하다.

Abstract Recently, various applications for image processing have been ported to the mobile environment due to the expansion of the image shooting on the mobile device. However, in the case of extracting the image foreground, which is one of the most important functions of image synthesis, is difficult since it needs complex calculation. In this paper, we propose an video synthesis technique that can divide images captured by mobile devices into foreground / background and combine them in real time on target images. Considering the characteristics of mobile shooting, our system can extract automatically foreground of input video that contains weak motion when shooting. Using SIMD and GPGPU-based acceleration algorithms, SD-quality images can be processed on mobile in real time.

Key Words : Foreground extraction, GPGPU, Image processing, Real-time solution, Smart application, Video matting

1. 서론

영상 합성 기술은 영상 콘텐츠 제작에 있어 반드시 사용되는 필수적인 기술로서, 원본 영상 내에 있는 영상 오브젝트를 목표 영상에 자연스럽게 붙여 넣는 기술을 뜻한다. 이러한 영상 합성을 진행하기 위해선 촬영된 영상속의 피사체만 추출하는 전경 (foreground) 추출이 필수적이며, 일반적으로는 크로마키(Chroma Key)와 같은 스튜디오에 제한된 접근방식이 사용되

고 있다. 이러한 크로마키 기술의 경우 촬영 단계에서부터 원본 영상의 합성을 고려하여 구성된 스튜디오에서만 사용 가능한 기술이기에 비전문적인 사용자의 사용 및 촬영 후 후반 작업에서는 사용이 힘들다는 단점을 가진다. 따라서 일반인 사용자들의 경우 후반작업으로 접근하기 위해선 비디오 매팅 (matting) 또는 세그멘테이션 (segmentation)과 같은 영상처리 기반의 기술이 사용되고 있다 [1,2,3].

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A3B03033656)

* Division of Computer, Media and Industrial Engineering, Kangwon National University. (media19@kangwon.ac.kr)

Received September 19, 2018

Revised October 01, 2018

Accepted October 15, 2018

이러한 후반적업 기반의 기술의 경우 촬영 시 영상 추출에 대한 고려가 필요하지 않다는 점에서 영상 촬영의 비용을 감소시킬 수 있는 장점을 가지며, 추가적으로 이미 제작되어있는 무한한 영상 데이터베이스 상에서 오브젝트를 추출할 수 있다는 장점을 가진다. 하지만 기존의 매칭 및 세그멘테이션 기반 기술의 경우 전경과 배경의 구분이 모호한 영역의 합성에서 문제점이 발생하며, 삼색맵 구성에 있어 사용자의 과도한 인터랙션 요구 및 높은 컴퓨팅 파워 등의 문제점으로 인한 실시간 처리가 불가능하다는 단점을 가지며, 이는 모바일 디바이스에 활용 불가능하다는 문제점을 야기시킨다.

본 연구에선 이러한 한계점을 극복하기 위한 모바일 기기를 위한 영상 합성 기술을 제안하고자 한다. 모바일 기기에 탑재된 카메라를 이용하여 촬영된 비디오 영상에 대해, 자동 혹은 사용자의 수동 보정에 기반한 삼색맵(Trimap)을 구성하고, 이를 자동으로 지오데식 거리맵 [4] 으로 변환하여 전경과 후경을 분리한 후, 영상 색상을 조정한 뒤 타 영상에 합성하는 순서로 진행된다 (그림 1 참조).

본 연구가 가지는 장점은 다음과 같다. 첫째, 흔들림에 강인한 전경 추적 기능을 사용하여, 사용자가 모바일 기기 상에서 손쉽게 사용할 수 있도록 최소한의 인터랙션만을 요구하는 기술은 제안한다. 둘째, 이미 지오데식 거리 기반의 가속화된 매칭 기술을 사용하여, 모바일 기기와 같은 제한된 컴퓨팅 환경에서도 수행이 가능하도록 한다. 제안된 기술의 경우 모바일 디바이스에 실시간 처리를 목표로 하였으며, 이를 해결하기 위해 Renderscript 및 Metal을 사용하여 최적화된 결과를 진행하였으며, 추가적으로 Neon을 이용한 SIMD 연산기반의 최적화 작업을 통해 SD영상에 대해 모바일상에서 실시간으로 처리가 가능한 시스템을 구성하였다.

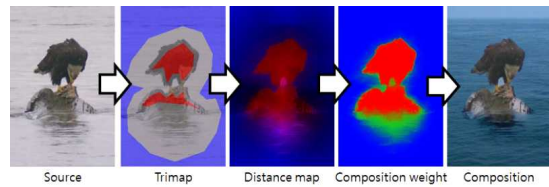


그림 1. 시스템 구성도
Fig. 1. System Overview

2. 자동 삼색맵 구성

사용자로부터 촬영 또는 입력된 영상의 전경을 자동 분리하기 위하여, 전후경 분리를 위한 세가지 영역, 즉 배경, 전경, 그리고 모호한 영역을 구분하는 삼색맵을 자동 구성하는 작업을 진행하여야 한다. 본 논문에서 사용한 모바일 영상 기반 삼색맵 계산 시스템은 그림 2와 같이, 비선형 전경추출 알고리즘의 대표적 방법인 ViBE를 모바일 환경에 맞게 개선하여 구현하였다.

2.1 ViBE기반의 전경확률맵 구성

최초로 사용자로부터 입력된 배경이미지 또는 배경만이 존재하는 영상으로 시작한다는 가정으로, 그림 3과 같이 블록기반의 SGM (single-Gaussian model) 형태의 배경정의자를 구성한다. 배경학습의 기본 전략은 ViBE [3] 알고리즘을 사용하였으며 스마트폰 카메라의 자동 노출 조절을 고려하여 Lab 컬러 공간에서 밝기 정보 L을 제외한 ab정보만을 기반으로 학습을 진행하였다. 또한 추후 들어오는 입력 영상에 대해 배경정의자에 어긋난 부분을 다시 SGM 형태의 전경정의자로 구성하여, Dual-SGM 모델을 생성한다. 입력되는 영상에 대해 배경 및 전경에 정의자와 모두 어긋나는 부분을 실제 전후경 영역이 분리될 모호영역으로 가정하여 삼색맵을 구성하였다.



그림 2. 영상의 자동 삼색맵 생성 시스템
Fig. 2. Automatically generation of trimap for video.

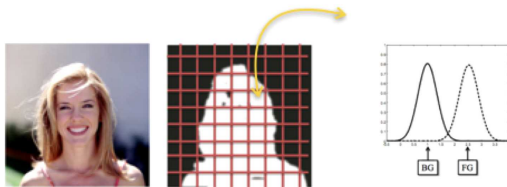


그림 3. 영상의 블록조각화 및 각 블록별 Dual SGM 생성.
Fig. 3. Block based Dual-SGM generation.

2.2 전경확률맵 개선 알고리즘

기존 ViBE [3] 알고리즘의 경우 전후경 분리를 위한 척도로 Lab 컬러공간에서 노출의 영향을 받는 L채널을 제외한 ab 색상 채널값을 고려한 전경과 배경의 차이를 계산하였다. 이러한 방식의 경우 자동 화이트 밸런스에 의해 변이된 L값에 대한 고려가 없어 안정적인 전후경 분리가 가능하다는 장점을 가지지만 반대로 색상값이 없는 무채색의 경우는 구분이 불가능한 문제점이 발생한다. 이러한 문제점을 해결하기 위해 본 연구에서는 사람에 인식에 기반한 색상차를 제공하는 CIEDE94 또는 CIEDE2000 [5] 기법을 적용하여 색상차 모델의 향상이 가능하다. CIEDE 2000기법의 경우 복잡한 계산량이라는 한계점으로 모바일 환경에서 적용이 힘들다는 문제점이 있었으며, 따라서 CIEDE94 기법을 적용하여 색상차 모델을 계산하는 모듈을 개발하였다 (그림 4 참조).

또한 Vibe의 한계점중 하나인 텍스처가 있는 영상에 대한 전후경 분리가 힘들다는 점을 보완하기 위하여 본 논문의 시스템에서는 단일픽셀 정보를 사용한 비교가 아닌 center-surround 기반의 Local Binary Similarity Patterns (LBSP)를 사용한 전후경 분리 기법 모듈을 시스템에 추가하여 적용하였다. LBSP기법의 경우 아래 수식과 같이 단일 픽셀을 중심으로 주변 픽셀들과의 화소 값을 비교하여 그 값이 크거나 같

면 1을, 그렇지 않으면 0으로 표현한다.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c) 2^p$$

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

이렇게 주변 픽셀들과의 비교를 통해 나타낸 2진 값 LBP(Local Binary Pattern)은 특정 픽셀을 중심으로한 색상의 정보 및 텍스처의 정보를 표현한다고 판단될 수 있다. LBP계산을 위해 어떤식으로 주변픽셀을 정의하느냐에 따라, 이진 패턴이 달라지며 본 시스템에서는 모바일 환경을 고려하여 16bit 기반의 2-ring구조의 이웃 픽셀을 정의하여 이진 패턴을 계산하였다. 최초 입력된 배경 이미지에 대한 각 픽셀별 형태 생성자를 LBP기반으로 생성하며, 이후에 들어오는 입력이미지에 대해 KLT 트래킹에 의한 호모그래피 안정화를 거친 결과물에 대해 LBP간 유사도를 분석한다. 만약 정해진 임계값 이내로 유사도가 판단되면 배경으로 판단하며 임계값을 벗어난 유사도를 보였을때는 전경으로 판단한다.

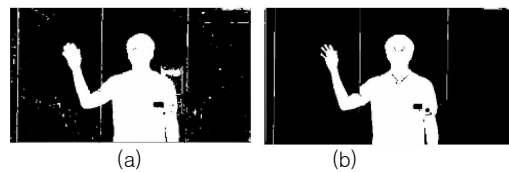


그림 4. 색상차 모델에 따른 전경확률맵 결과 차이:(a) Lab 기반의 색상차 모델; (b) CIEDE94기반 색상차 모델.
Fig. 4. Foreground map calculation based on different color model: (a) Lab-based result; (b) CIEDE94-based result.

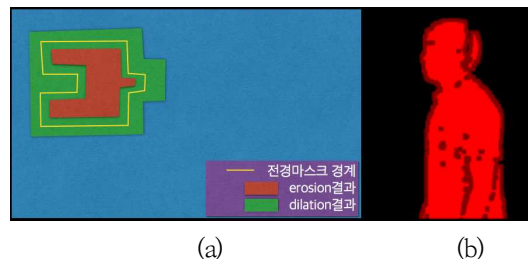


그림 5. 모폴로지 연산 기반의 삼색맵 구성: (a) 삼색맵 구성을 위한 모폴로지 적용 구성 (파란색:배경, 빨간색:전

경, 초록색:모호영역); (b) 실제 삼색맵 구성 결과.
 Fig. 5. Trimap generation using morphology operation s: (a) result of morphology operations (blue: background, red: foreground, green: context region); (b) trimap example.

Vibe알고리즘을 사용한 전경 추출의 경우, 흔들리는 카메라에서도 높은 정확도를 보이는 장점을 가지나, 노이즈에 민감하며 잘못된 단일 픽셀형태의 잘못된 영역구분이 생기는 문제점이 발생한다. 이를 해결하기 위하여 모폴로지 연산 [6] 을 사용하여 완전한 전경과 완전한 배경 사이에 존재하는 모호한 영역에 대한 연결성을 보정하는 연산을 추가하였다. 우선 최초로 구해진 전경 마스크에 대해 erosion연산을 진행하여 바운드리를 잘라내고 불필요한 노이즈를 제거하였으며, 이 결과에 대해 dilation 연산을 보다 큰 엘리먼트로 진행하여 두 영역에 차이를 모호한 영역으로 구성하였다 (그림 5 참조).

3. 지오데식 거리기반의 가속화된 매팅

지오데식 거리 (Geodesic distance)란, 기하학적 용어로서 임의의 곡면상 두 점 사이에 곡면 위를 지나가는 최단 거리를 의미한다. 이미지 지오데식 거리는 위 지오데식 거리를 이미지에 이용한 것으로, 이미지를 가로축/세로축/색상축으로 이루어진 고차원 곡면으로 보았을 때, 이 곡면 상에 위치한 두 픽셀 간의 거리를 의미한다. 임의의 이미지에 대하여, 이미지 지오데식 거리맵은 각각의 픽셀로부터 마스크까지의 지오데식 거리 중 최소값을 각 픽셀에 저장한 것을 의미하며, 이를 계산하는 과정을 이미지 지오데식 거리 트랜스폼이라 한다. 최근 이러한 지오데식 거리 트랜스폼을 활용하여 영상의 매팅을 진행하는 연구가 발표 [4] 되었으며, 본 연구에서는 이 알고리즘을 SIMD 및 GPGPU 기반으로 가속화 하기 위해 Android에서 동작하는 Renderscript를, 애플은 iOS에서 동작하는 Metal을 사용하여 구현하였다.

이미지 지오데식 거리 계산 알고리즘은 이론상 복잡도가 픽셀의 개수에 선형비례하는 빠른 알고리즘이지만, 모바일 기기 등 한정적인 계산 자원을 갖는 장치에서 수행하기 위해서는 추가적인 최적화가 필요하다.

구현과정에서 지오데식 거리 계산 알고리즘 상에서 이웃하는 픽셀 간의 차이값을 계산하는 부분에서 두 번의 스캔라인 과정을 수행하는 동안 중복되는 것을 확인할 수 있었으며, 이웃 픽셀간 차이를 저장하는 공간을 추가적으로 확보하여 미리 계산하는 형태로 최적화를 진행하였다. 이를 통해 삼색맵 구성 및 매팅까지의 시간을 iPhone 7를 사용했을때 1frame당 40 msec의 속도를 가져왔으며 실시간 처리가 가능한 어플리케이션의 개발을 완성하였다 (그림 6 참조).

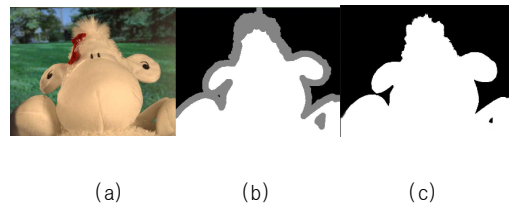


그림 6. 지오데식 거리기반 매팅 결과: (a) 입력영상; (b) 삼색맵 구성; (c) 전경 알파맵 결과.
 Fig. 6. Result of geodesic distance-based matting: (a) input video; (b) trimap generation; (c) matting result.

4. 색온도 기반의 색보정 알고리즘

제한된 시스템의 경우 분리된 전경을 타 배경에 합성하는 것을 최종 목표로 한다. 하지만 이러한 합성에 있어 이질감을 최소화시키기 위해서는 추출된 전경과 새로운 배경간의 색보정 작업이 필수적이다. 하지만 전경 분리에서 소모된 계산에 추가적으로 기존 색보정 알고리즘을 모바일 환경으로 옮기는 것은 본 시스템의 목표인 실시간 처리에는 불가능한 요소로 판단되었다. 따라서 최소한의 계산으로 최대한의 효율을 얻기위해 본 시스템에서는 영상의 색온도를 바탕으로한 단순하지만 효율적인 색보정 알고리즘을 차용하였다.

영상의 색온도 변화를 위해 다음과 같은 세가지 가정에 기반하여 색상보정알고리즘을 구성하였다: 1) 색온도가 6600K 이상에서는 Red는 항상 255값을 가진다; 2) 색온도가 2000K 이하에서는 Blue값은 항상 0이다; 3) 색온도가 6500K 이상에서는 Blue값은 항상 255이다. 이러한 3 가지 가정에 벗어나는 색온도의 경우는 CIE 1964 [7] 에 기반한 색온도 별 색상 감쇠 커브를 사용하여 각 컬러공간 별 변화를 진행하였으

며, 합성될 전후경의 색온도를 동일시하여 단순하지만 효과적인 색상 보정기능을 추가하였다 (그림 7 참조).

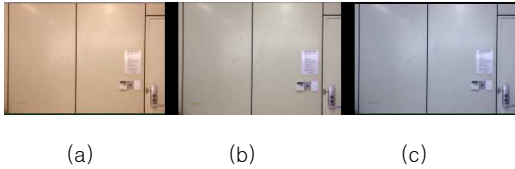


그림 7. 영상의 색온도 조절에 의한 배경 톤 변화: (a) 5800K로 색온도 변화 배경; (b) 6200K의 원본 배경; (c) 6800K로 색온도 상승배경.

Fig. 7. Color tone changing using color temperature: (a) background of 5800K color temperature; (b) background of 6200K color temperature; (c) background of 6800K color temperature.

5. 실험결과 및 분석

본 연구의 결과는 그림 8에 나타나 있다. 가속화된 구현을 통해 모바일 환경에서 촬영된 영상에 대해 모바일 기기에서 실시간으로 전경 추출 및 배경 합성이 가능한 어플리케이션을 개발하였으며 성능을 검증하기 위해 속도, 정확도, 시간연속성의 3가지 기능의 성능을 측정해 보았다.

우선 속도의 경우 모바일 환경에서 주로 사용되는 720P 해상도의 영상에 대해 범용화된 모바일 기기인 iPhone 7의 성능으로 모든 연산을 100.75 msec. 에 처리가능함을 보였다. 실시간을 보충하기 위해선 1초에 40 msec.가 보장되어야 하나, 최근 모바일 디바이스의 성능향상에 기인했을 때 실시간의 근접한 성능으로 판단된다.

전경 추출의 정확성을 증명하기 위해 알파매핑 성능 비교 사이트 [8] 에서 제공하는 테스트 이미지 셋에 대해 알파맵을 계산하고, 이를 기준 알파맵과 절대차합(Sum of Absolute Difference, SAD)를 비율 형태로 계산한 결과 오차율이 2.359%로 높은 수준의 결과를 보였으며 이는 기존 매핑 기법에 비해 낮지 않는 수준의 결과를 모바일에서 보였다.

비디오의 경우 시간연속성이 유지되어야 하므로, 이를 기존 기법과 비교하기 위해 알파값이 존재하는 기존 영상과 본 시스템에서 추출된 영상에 대한 임의의

위치에 100 pixel에대한 시간별 알파값의 표준편차를 측정한 결과, 약 0.08의 평균편차의 차이를 보였으며, 이는 충분한 정도의 시간연속성을 가진다고 판단하였다.

마지막으로 본 연구에서 제안한 향상된 ViBE알고리즘을 사용한 결과 검증을 진행하였다. 그림 9와 같이 기존 ViBE알고리즘을 사용한 결과보다 전경추출 정확도를 높일수 있었으며, 특히 배경에서 텍스처가 다른 물체가 존재하는 영역에서 보다 안정적인 결과를 가져왔음을 증명하였다.

6. 결론

본 논문에서는 모바일에서 촬영된 영상에 대해 자동으로 삼색맵을 구성하여, 그 내용을 바탕으로 전경을 분리하는 어플리케이션 개발 방법을 소개하였다. 최소한의 연산을 가지는 방법을 선택하여 이를 가속화 하여 구현함으로써, 모바일 기기에서도 실시간으로 영상 합성이 가능한 결과를 보였다는 점에서 의미를 가진다 판단된다.

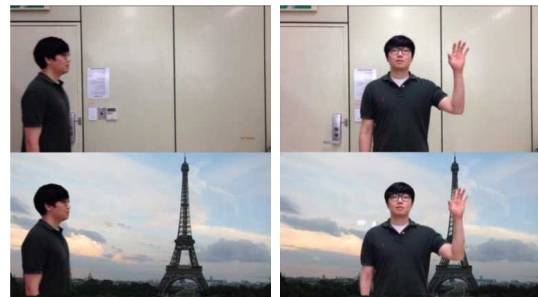
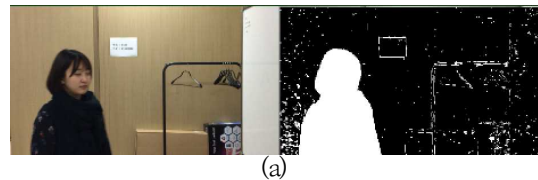
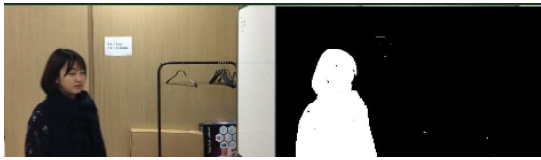


그림 8. 모바일 기반 영상의 전경 추출 및 합성 결과
Fig. 8. Results of foreground extraction and composition



(a)



(b)

그림 9. 향상된 ViBE알고리즘으로 인한 결과 변화: (a) 기존 ViBE 알고리즘 기반 전경추출; (b) 본 논문의 향상된 ViBE알고리즘 기반 전경추출

Fig. 9. Results of enhanced ViBE algorithm: (a) foreground alpha map by ViBE algorithm; (b) foreground alpha map by our enhanced algorithm.

본 논문이 제안하는 시스템의 한계점은, 비록 Vibe 알고리즘이 가벼운 카메라의 움직임에는 영향을 크게 받지 않지만 과도한 카메라의 움직임이 있을 때 삼색 맵 구성에 있어 잘못된 결과를 가져 올수 있다는 점이다. 따라서 가속화된 트래킹 알고리즘을 추가하여 보다 안정적인 삼색 맵 구성을 가져왔을 때, 활용도 높은 어플리케이션으로 사용 될 수 있을것으로 예상된다.

REFERENCES

- [1] R. Cucchiara, C. Grana, M. Piccardi, A. Prati: "Detecting moving objects, ghosts, and shadows in video streams", IEEE Trans. Pattern Anal. Mach. Intell., 25 (10) (2003), pp. 1337-1342
- [2] Li Yao and M. Ling: "An improved mixture-of-Gaussians background model with frame difference and blob tracking in video stream" Sci. World J. (2014) (2014), pp. 1-9
- [3] Olivier Barnich and Marc V. Droogenbroeck, "ViBe : A universal background subtraction algorithm for video sequences." IEEE Transactions on Image Processing, 20(6) :1709-1724,
- [4] Xue Bai and Guillermo Sapiro, "A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting", International Journal of Comput Vision, pp 113-132, 2009
- [5] Yongjin Kim , Cheolhun Jang, Julien Demouth , Seungyong Lee: "Robust color-to-gray via non linear global mapping", ACM Transactions on Graphics (TOG), v.28 n.5, 2009
- [6] Rafael C. Gonzalez and Richard E. Woods: Digital Image Processing(4th edition), 2017
- [7] Wyszeccki and Günther. "Proposal for a New Color-Difference Formula". Journal of the Optical Society of America. JOSA. 53 (11): 1318-1319. 1963

[8] www.alphamatting.com

저자약력

윤종철 (Jong-Chul Yoon)

[정회원]



<관심분야>

- 2003 아주대학교 미디어학(학사)
- 2005 아주대학교 미디어학(석사)
- 2010 연세대학교 컴퓨터과학(박사)
- 2010~현재 강원대학교 컴퓨터 미디어산업공학부 조교수

컴퓨터 그래픽스., 영상처리, 정보시각화