

특집논문 (Special Paper)

방송공학회논문지 제20권 제3호, 2015년 5월 (JBE Vol. 20, No. 3, May 2015)

http://dx.doi.org/10.5909/JBE.2015.20.3.414

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

배경 영역의 시간적 일관성이 향상된 고해상도 깊이 동영상 생성 방법

신 동 원^{a)}, 호 요 성^{a)‡}

Temporally-Consistent High-Resolution Depth Video Generation in Background Region

Dong-Won Shin^{a)} and Yo-Sung Ho^{a)‡}

요 약

3차원 영상 시스템에서 깊이 영상은 3차원 콘텐츠를 표현하는데 있어 매우 중요한 역할을 수행한다. 그러나 깊이 카메라로부터 얻은 원본 깊이 영상은 해상도가 색상 영상에 비해 매우 작고 시간적 흐름의 측면에서 관찰하였을 때 깊이 값이 불안정하게 진동하는 깜빡임 문제가 발생한다. 이 문제는 시청자들이 3차원 콘텐츠를 감상할 때 불편한 느낌을 초래한다. 이 논문에서는 원본 깊이 영상의 저해상도 문제를 해결하기 위해 3차원 워핑과 깊이 가중치가 추가된 결합형 양방향 업샘플링 방법을 사용한다. 다음으로 깊이 영상의 배경 영역에서 발생하는 깜빡임 문제를 해결하기 위해 전경과 배경을 분리한 뒤, 전경 영역에는 업샘플링된 깊이 영상을 사용하고 배경 영역에는 시간적 평균값 필터 영상을 이용했다. 실험결과를 제안하는 방법이 시간적 일관성이 향상된 고해상도의 깊이 영상을 생성함을 보였다.

Abstract

The quality of depth images is important in the 3D video system to represent complete 3D contents. However, the original depth image from a depth camera has a low resolution and a flickering problem which shows vibrating depth values in terms of temporal meaning. This problem causes an uncomfortable feeling when we look 3D contents. In order to solve a low resolution problem, we employ 3D warping and a depth weighted joint bilateral filter. A temporal mean filter can be applied to solve the flickering problem while we encounter a residual spectrum problem in the depth image. Thus, after classifying foreground and background regions, we use an upsampled depth image for a foreground region and temporal mean image for background region. Test results shows that the proposed method generates a time consistent depth video with a high resolution.

Keyword : Kinect v2, depth flickering, temporal consistency, upsampling, warping

a) 광주과학기술원 정보통신공학부(School of Information and Communication)

‡ Corresponding Author : 호요성(Yo-Sung Ho)

E-mail: hoyo@gist.ac.kr

Tel: +82-62-715-2258

ORCID: http://orcid.org/0000-0002-7220-1034

※ 이 논문의 연구결과 중 일부는 "IWAIT 2015"에서 발표한 바 있음.

· Manuscript received March 17, 2015; revised May 6, 2015; accepted May 12, 2015.

1. 서론

최근 일반 소비자들의 3차원 영상에 대한 관심이 증가함에 따라 우리는 3차원 영상 기술이 접목된 다양한 제품에 쉽게 접근할 수 있게 되었다. 기존의 영화관에 3D 영화 상

영 시설을 설치하거나 시장에 출시된 많은 수의 3DTV를 통해 3차원 콘텐츠를 쉽게 볼 수 있는 환경이 제공되었다. 이러한 장비를 이용하여 우리는 2차원 영상에 존재 하지 않는 실감나는 깊이감을 느낄 수 있게 되었다.

이러한 깊이감은 다양한 깊이 획득 방법을 통해 얻어진 깊이 영상으로부터 느낄 수 있다. 깊이 영상을 획득하는 방법은 크게 능동형과 수동형의 2가지로 분류된다. 능동형 방식은 3차원 장면에 적외선이나 초음파와 같은 물리적인 광선을 직접적으로 투사하는 깊이 카메라를 이용하여 3차원 정보를 획득하고 수동형 방식은 장면에 대한 스테레오 영상으로부터 텍스처 패턴 매칭을 수행하여 3차원 정보를 획득한다^[1]. 만약 깊이 영상이 정확하게 획득되지 않는다면, 이 단계에서 발생한 잡음이 사용자단까지 전파되어 시청자는 완성된 3차원 콘텐츠에 대해 불편한 시청감을 느끼게 될 것이다. 따라서 깊이 영상은 전반적인 3차원 영상 시스템에 있어 중요한 역할을 수행한다고 할 수 있다^[2].

이 논문에서는 3차원 영상 연구 분야에서 널리 쓰이는 마이크로소프트 키넥트 버전 2에서 발생하는 깊이 영상의 문제점을 살펴보았다. 그림 1은 키넥트 버전 2의 외관을 보여주어 주고 있고, 표 1은 마이크로소프트 키넥트 버전 1과 버전



그림 1. 마이크로소프트 키넥트 버전 2의 외관
 Fig 1. Appearance of Microsoft Kinect V2

표 1. 마이크로소프트 키넥트 버전 1과 버전 2의 비교
 Table 1. Comparison of Microsoft Kinect V1 and Kinect V2

	Kinect V1	Kinect V2
Depth acquisition method	Structured Light	Time of Flight (ToF)
Color resolution	640×480	1920×1080
Depth resolution	320×240	512×424
Depth range	0.5-4.0 meters	0.5-4.5 meters

2를 비교한다^[3].

키넥트 버전 2의 성능은 키넥트 버전 1과 비교해 보았을 때 해상도와 깊이 범위 면에서 상당한 향상이 이루어졌으며 구조광 방식에서 ToF 방식으로 바뀌면서 깊이 값의 정확도가 향상되었고 깊이 데이터 손실이 감소했다.

그러나 이러한 향상에도 불구하고 키넥트 버전 2의 깊이 카메라로부터 얻은 원본 깊이 영상은 시간적인 측면에서 관찰하였을 때 불안정한 모습을 보인다. 즉, 깊이 값이 시간 흐름에 따라 크게 흔들리는 깜빡임 문제가 발생하는데 이것은 일반적으로 깊이 카메라를 사용하는 능동형 깊이 획득 방식에서 많이 발생한다. 그림 2는 마이크로소프트 키넥트 버전 2에서 획득한 깊이 동영상을 보여준다.



그림 2. 키넥트 버전 2에서 획득한 깊이 동영상
 Fig 2. Depth video acquired from Kienct V2

이 깊이 동영상에서 주로 물체의 경계를 따라 깊이 값이 나타났다가 사라지는 깜빡임 현상이 나타나는 것을 확인할 수 있다. 그림 3은 깜빡임 현상이 나타나는 영역을 확대하여 보여주고 있다.

확대된 깜빡임 영역을 살펴보면 영상의 좌상단 영역과 책상의 경계부분 등에서 일관적이지 않은 깊이 값의 양상을 확인할 수 있다. 결과적으로 이러한 형태의 깜빡임 현상이 시청자들에게 어지럽고 불편한 3차원 콘텐츠를 제공하



그림 3. 확대된 깜빡임 영역
Fig 3. Enlarged flickering region

도록 만든다.

이 논문에서는 깜빡임 문제를 제거하기 위한 방법에 대해 소개한다. 2장에서는 제안하는 깜빡임 제거 방법의 흐름도와 구현방법에 대한 자세한 내용을 다룬다. 3장에서는 제안하는 방법으로부터 얻은 결과와 원본 영상을 비교하는 실험에 대해 논의하고 4장에서 결론으로 마무리 한다.

II. 제안하는 깜빡임 제거 방법

그림 4는 제안하는 방법의 흐름도를 보여준다^[4].

먼저 키넥트 버전 2로부터 얻은 색상 영상과 깊이 영상의

시점과 해상도를 일치시키기 위해 카메라 파라미터를 얻는 과정이 필요하다. 이 논문에서는 칼텍에서 제공하는 매트랩 카메라 캘리브레이션 툴박스를 이용하여 색상영상과 깊이영상의 카메라 캘리브레이션 과정을 수행하여 카메라 파라미터를 획득했다^[5]. 이 카메라 파라미터를 이용하여 3차원 워핑을 수행하면 깊이 카메라 시점의 깊이 영상을 색상 카메라의 시점과 일치시킬 수 있다^[6]. 동시에 색상 영상의 해상도와 일치하는 희소한 형태의 깊이 영상을 얻을 수 있다. 다음으로 깊이 값이 없는 부분에 적절한 깊이 값을 채우기 위해 깊이 가중치가 추가된 결합형 양방향 필터를 수행한다. 그리고 이 결과로 나온 업샘플링된 깊이 영상을 현재 프레임으로부터 특정한 개수만큼 배열에 저장하고 이 배열

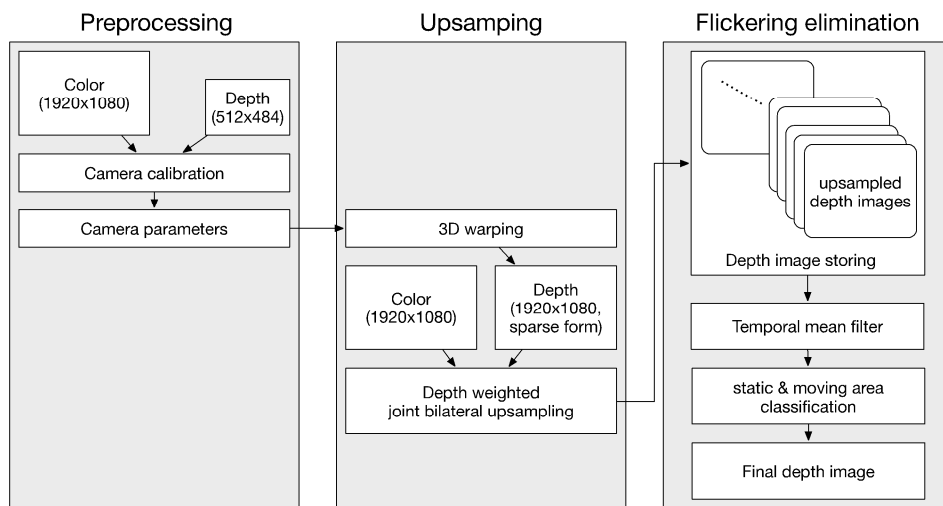


그림 4. 제안하는 깜빡임 제거 방법의 흐름도
Fig 4. Flowchart of the proposed flickering elimination

에 시간적 평균 필터를 수행하여 깜빡임을 제거한다. 그런데 시간적 평균 필터를 적용하면 깜빡임 문제는 사라지지만 이전 깊이 프레임의 잔상이 남는다는 문제가 발생한다. 따라서 이러한 잔상 문제를 해결하기 위해 깊이 영상을 정적 영역과 동적 영역으로 분류하고 영역의 특성에 따라 시간적 평균 깊이 영상과 업샘플된 깊이 영상을 적응적으로 사용하도록 했다.

1. 깊이 가중치가 추가된 결합형 양방향 필터

이 논문에서는 3차원 워핑을 통해 색상 영상과 깊이 영상의 시점과 해상도를 일치시킨 다음, 이 단계로부터 획득한 희소한 형태의 깊이 영상에서 깊이 값이 없는 부분에 적절한 깊이 값을 채워 넣기 위해 깊이 가중치가 추가된 결합형 양방향 필터를 제안한다. 그림 5는 희소한 형태의 깊이 영상의 예를 보여준다.

그림 5의 왼쪽에 확대된 영역 (1)과 영역 (2)를 살펴보면 화소의 모든 위치에 대해 깊이 값이 채워져 있지 않고 희소하게 깊이 값이 분포되어 있는 것을 확인할 수 있다. 이는 저해상도 깊이 영상의 각 화소들이 고해상도의 색상 영상의 크기에 맞게 워핑되면서 발생하는 문제이다. 또한 그림 5의 오른쪽에 확대된 영역 (3)과 영역 (4)를 살펴보면 깊이 값이 좀 더 밀집되어 있는 부분을 관찰할 수 있다. 이 부분 또한 워핑에 의해 깊이 카메라 시점이 색상 카메라 시점으

로 이동하면서 전경에 대한 깊이 영역과 배경에 대한 깊이 영역이 겹쳐지면서 발생하는 폐색영역 문제이다.

3차원 워핑에 의해 발생하는 이러한 문제점들을 해결하기 위해 이 논문에서는 기존의 결합형 양방향 필터^[7]에 깊이 가중치가 추가된 형태를 제안한다. 식 (1)은 제안하는 방법의 계산 구성을 나타낸다.

$$D_o(x, y) = \frac{\sum_{u \in \vec{u}} \sum_{v \in \vec{v}} W(u, v) \cdot D_i(u, v)}{\sum_{u \in \vec{u}} \sum_{v \in \vec{v}} W(u, v)} \quad (1)$$

$$(\vec{u} = \{x-r, \dots, x+r\}, \vec{v} = \{y-r, \dots, y+r\})$$

D_o 는 출력 깊이 화소를 의미하고 D_i 는 입력 깊이 화소를 나타낸다. (x, y) 와 (u, v) 는 각각 필터 커널에서 중심 화소의 위치와 이웃 화소의 위치를 나타낸다. 벡터 \vec{v} 와 벡터 \vec{u} 는 각각 필터 커널에서 수평 위치 집합과 수직 위치 집합을 의미한다. 마지막으로 W 는 가중치 인자를 나타낸다. 기존의 결합형 양방향 필터의 가중치 인자는 2개의 가우시안 함수 곱으로 구성되어 있다. 이 가우시안 함수는 중심 위치에서 가장 높은 값을 가지고 양 끝으로 갈수록 0에 가까운 값을 가지는 종모양의 함수이다. 우리는 기존의 가중치 인자에 중심 화소와 이웃 화소의 깊이 화소 값 차이를 반영하는 가중치 인자를 추가함으로써 3차원 워핑으로부터 발생



그림 5. 희소한 형태의 깊이 영상
 Fig 5. Depth image in a sparse form

하는 문제를 해결한다. 식 (2)는 제안하는 가중치 인자 W 를 나타낸다.

$$W(u,v) = \begin{cases} 0 & \text{if } D_i(u,v) = 0 \\ g(u,v) \cdot f(u,v) \cdot d(u,v) & \text{otherwise} \end{cases} \quad (2)$$

$g(u,v)$ 와 $f(u,v)$ 는 각각 필터 커널의 중심 화소와 이웃 화소간의 색상 차이와 거리 차이를 반영하는 가중치이다. 여기까지가 기존의 가중치이고 우리는 여기에 중심화소와 이웃 화소간의 깊이 차이를 반영하는 가중치 $d(u,v)$ 를 추가적으로 사용한다. 식 (3), 식(4), 식(5)는 각 가중치 인자의 구성을 나타낸다.

$$g(u_x, v_y) = \exp \left\{ -\frac{|I(x,y) - I(u_x, v_y)|^2}{2\sigma_g^2} \right\} \quad (3)$$

$$f(u_x, v_y) = \exp \left\{ -\frac{(x-u_x)^2 + (y-v_y)^2}{2\sigma_f^2} \right\} \quad (4)$$

$$d(u_x, v_y) = \exp \left\{ -\frac{|D_i(x,y) - D_i(u_x, v_y)|^2}{2\sigma_d^2} \right\} \quad (5)$$

우리는 이러한 깊이 가중치가 추가된 결합형 양방향 필터를 이용하여 3차원 워핑으로부터 발생하는 문제들을 해결할 수 있다. 그림 6은 기존의 결합형 양방향 필터와 제안하는 필터의 결과를 비교한 그림이다.

두 방법 모두 희소한 형태의 깊이 영상을 적절한 깊이 값으로 채우는 것은 성공했다. 하지만 전경과 배경의 깊이 값이 겹치면서 발생하는 폐색영역(영상에서 손과 얼굴의 오른쪽 부분)에 대해서는 제안하는 방법만이 깊이 값을 잘 채워주는 것을 확인할 수 있다.

2. 시간적 평균 필터

이 절에서는 앞에서 제안한 방법으로 얻은 업샘플된 깊이 영상을 이용하여 배경 영역에 대한 깜빡임 문제를 해결



(a) 기존의 결합형 양방향 필터



(b) 제안하는 필터

그림 6. 기존의 방법과 제안하는 방법의 결과 비교
Fig 6. Comparison with a convention and a proposed method

하는 것에 대해 살펴본다. 시간적 평균 필터는 깊이 영상의 시간적 배열에 대해 평균값을 취하는 필터인데 식 (6)은 이러한 시간적 평균 필터에 대한 구조를 나타낸다.

$$D_M(x,y) = \frac{\sum_{t=1}^N D_o^t(x,y)}{N} \quad (6)$$

$D_o^t(x,y)$ 는 t 프레임에서 (x,y) 위치의 업샘플된 깊이 화소를 의미하고 N 은 영상 배열안에 저장된 깊이 영상의 개수를 의미한다. 즉, 이 식은 영상 배열에 저장된 모든 깊이 영상에 대해서 (x,y) 위치에서의 평균값을 얻는 것을 의미한다. 깜빡임 현상은 제거할 수 있지만 이전 깊이 영상이 다음 깊이 영상에 영향을 미치는 잔상 문제를 만나게 된다. 이 논문에서는 이러한 잔상 문제를 해결하기 위해 깊이 영상을 정적 그리고 동적 영역으로 분류하고 영역의 종류에 따라 적응적으로 처리한다.

3. 동적, 정적 영역 분류

주로 정적 영역은 움직임이 없는 배경 영역을 의미하고 동적 영역은 물체의 움직임이 많은 전경영역을 의미한다. 먼저 영상 배열에 대해 전경에 있는 물체의 움직임을 기록하기 위해 식 (7)을 수행한다.

$$D_{MT}(x, y) = \max_{0 \leq t \leq N} (D'_o(x, y)) \quad (7)$$

이 식은 영상 배열에 있는 모든 깊이 영상에 대해 (x,y) 위치의 최대값을 취한다.

움직임을 기록한 영상에서 깊이 문턱값을 이용하여 문턱값 이상은 동적 영역으로, 문턱값 이하는 정적 영역을 분리한 다음, 동적 영역에는 업샘플된 깊이 영상을 사용하고 정적 영역에는 시간적 평균 영상을 사용하여 배경 영상에서의 깜빡임을 제거했다.

III. 실험 결과

이 논문의 실험에서, 마이크로소프트 키넥트 버전 2가 사용되었다. 사용된 색상 영상의 해상도는 1920×1080이며 깊이 영상의 해상도는 512×424이다. 계산 속도를 가속화하기 위해 GPU(Graphics processing unit)를 이용한 병렬 계산을 수행하였고 사용된 GPU 모델은 nVidia Quadro K4000이다^[8].

정량적인 분석을 위해서 그림 8, 그림 9, 그림 10에 배경 영역에 대한 깊이 화소 값의 변화 추이를 나타내었다. 특별히 배경 영역에서 물체의 경계 부분의 깊이 화소 값에 대한 변화를 그래프로 나타내었고 깊이 값을 추출한 위치를 그림 7에서 시각적으로 표현했다. 수평축은 시간에 따른 프레임 수를 나타내고 수직축은 깊이 값을 의미한다. 파란색 선은 원본 깊이 동영상에서의 깊이 값의 변화를 나타내고 빨간색 선은 결과 깊이 동영상에서의 깊이 값의 변화를 나타낸다. 결과 그래프를 통해서 원본 깊이 동영상의 깊이 값이 심하게 진동하는 반면 결과 깊이 동영상의 깊이 값은 상대적으로 안정적인 모습을 보이는 것을 확인했다.



그림 7. 시간의 흐름에 따른 깊이 값 추출
Fig 7. Depth value extraction along with temporal flow

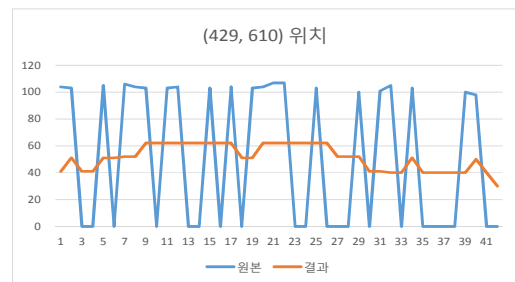


그림 8. (429, 610) 위치에서의 깊이 값 변화 그래프
Fig 8. Depth value graph at (429, 610)

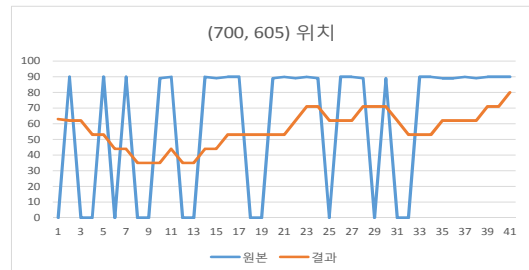


그림 9. (700, 605) 위치에서의 깊이 값 변화 그래프
Fig 9. Depth value graph at (700, 605)

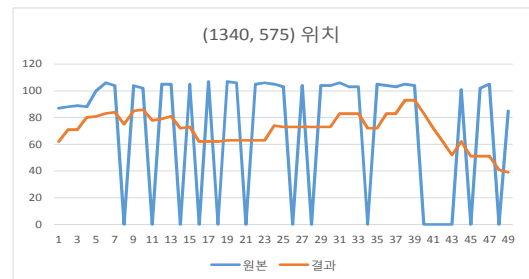


그림 10. (1340, 575) 위치에서의 깊이 값 변화 그래프
Fig 10. Depth value graph at (1340, 575)

IV. 결 론

이 논문에서는 마이크로소프트 키넥트 버전 2 깊이 카메라로부터 얻은 깊이 동영상에서 배경 영역에 대한 시간적 일관성을 향상하는 방법을 논의했다. 먼저 원본 깊이 영상과 색상 영상의 시점을 일치시키기 위해 3차원 워핑을 수행했고 해상도를 일치시키기 위해 깊이 가중값이 추가된 결합형 양방향 필터를 수행했다. 또한, 깊이 영상에서 발생하는 깜빡임 문제를 해결하기 위해 시간적 평균 필터를 이용했다. 그런데 이로부터 추가적으로 발생하는 잔상 문제를 해결하기 위해 정적 영역과 동적 영역을 추출하고 영역의 특성에 따라 동적 영역에서는 업샘플된 깊이 영상을 사용하고 정적 영역에서는 시간적 평균 영상을 사용함으로써 배경영역에 대한 깜빡임 문제를 해결했다. 영상의 특정 위치에서 깊이 값을 추출하여 그래프를 만드는 실험 결과를 통해 제안하는 방법이 원본 영상의 배경영역에 대해 깊이 값의 분산 정도를 평균적으로 17배 감소시키면서 깜빡임 현상을 현저하게 줄이는 것을 확인했다.

참 고 문 헌 (References)

- [1] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Image Commun.*, vol. 22, no. 2, pp. 217-234, Feb. 2007.
- [2] C. Fehn, "Depth-image-based rendering (DIBR), Compression and Transmission for a New Approach on 3-D TV," *Proc. of SPIE Conference Stereoscopic Displays and Virtual Reality Systems*, vol. 5291, pp. 93-104, Jan. 2004.
- [3] <http://www.microsoft.com/en-us/kinectforwindows/meet-kinect/features.aspx>
- [4] D. Shin and Y. Ho, "Flickering Elimination Method for Background Region in Depth Video," *Proc. of International Workshop on Advanced Image Technology*, OS10-73, Jan. 2015.
- [5] Camera Calibration Toolbox for MATLAB: <http://www.vision.caltech.edu/bouguetj>
- [6] W. R. Mark, L. McMillan, and G. Bishop, "Post-rendering 3D Warping," *Proc. of Symposium on Interactive 3D Graphics*, pp. 7-16, April 1997.
- [7] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 1-5, July 2007.
- [8] D. Shin and Y. Ho, "Real-time Depth Image Refinement using Joint Bilateral Filter," *Proc. of 2013 Korean Society of Broadcast Engineers Autumn Conference*, vol. 19, pp. 116-119, Nov. 2013.

저 자 소 개



신 동 원

- 2013년 : 금오공과대학교 컴퓨터공학과 졸업(학사)
- 2015년 : 광주과학기술원 정보통신공학부 졸업(석사)
- 2015년 ~ 현재 : 광주과학기술원 정보통신공학부 박사과정
- ORCID : <http://orcid.org/0000-0002-2769-3957>
- 주관심분야 : 3D 영상처리, 실감방송



호 요 성

- 1981년 : 서울대학교 공과대학 전자공학과 졸업(학사)
- 1983년 : 서울대학교 대학원 전자공학과 졸업(석사)
- 1989년 : Univ. of California, Santa Barbara, Dept. of Electrical and Computer Engineering.(박사)
- 1983년 ~ 1995년 : 한국전자통신연구소 선임연구원
- 1990년 ~ 1993년 : 미국 Philips 연구소 Senior Research Member
- 1995년 ~ 현재 : 광주과학기술원 정보통신공학부 교수
- ORCID : <http://orcid.org/0000-0002-7220-1034>
- 주관심분야 : 디지털 신호처리, 영상 신호 처리 및 압축, 디지털 TV와 고선명 TV, 멀티미디어 시스템, MPEG 표준, 3차원 TV, 실감방송