

원격 영상회의 시스템을 위한 깊이 영상 후처리 기술

정회원 이상범*, 종신회원 호요성*

Depth Video Post-processing for Immersive Teleconference

Sang-Beom Lee* *Regular Member*, Yo-Sung Ho* *Lifelong Member*

요 약

본 논문에서는 IPTV 환경의 원격 영상회의에서 화자 간의 자연스러운 시선 맞춤(eye contact)을 위한 깊이영상의 후처리 필터링 기술을 제안한다. 제안하는 방법은 깊이탐색 기술과 영상합성 기술을 사용해서 화자의 정면시점 영상을 합성한다. 하지만, 깊이영상을 탐색하는 과정에서 객체의 경계 불일치, 시간적 상관도 저하 등의 문제가 발생하기 때문에 이를 해결하기 위해 시간축으로 확장된 결합형 양방향 필터(joint bilateral filter)를 제안한다. 실험 결과를 통해, 제안하는 깊이영상의 후처리 필터링 기술이 정면시점 합성영상의 화질을 향상시켰고, 원격의 화자와 시선 맞춤이 가능한 것을 확인했다.

Key Words : Gaze correction, telepresence, joint bilateral filter, IPTV, three-dimensional video

ABSTRACT

In this paper, we present an immersive videoconferencing system that enables gaze correction between users in the internet protocol TV (IPTV) environment. The proposed system synthesizes the gaze corrected images using the depth estimation and the virtual view synthesis algorithms as one of the most important techniques of 3D video system. The conventional processes, however, causes several problems, especially temporal inconsistency of a depth video. This problem leads to flickering artifacts discomfoting viewers. Therefore, in order to reduce the temporal inconsistency problem, we exploit the joint bilateral filter which is extended to the temporal domain. In addition, we apply an outlier reduction operation in the temporal domain. From experimental results, we have verified that the proposed system is sufficient to generate the natural gaze-corrected image and realize immersive videoconferencing.

I. 서 론

원격 영상회의는 여러 가지 단말과 회선으로 이루어지는 통신시스템과 카메라, 디스플레이 장치, 마이크, 스피커를 이용해서 멀리 떨어져 있는 사람들의 모습을 보면서 마치 동일한 공간에서 함께 회의를 진행하는 느낌을 제공한다¹⁾. 최근에는 컴퓨터로도 회의가 가능하며, 많은 연구 기관에 의해 고화

질 및 고음질의 영상회의를 위한 기술들이 각광받고 있다²⁻⁴⁾. 또한, 파일을 공유하면서 회의를 진행할 수도 있고, 쌍방의 모니터 화면에 보조 장치로써 전자펜으로 글씨나 그림을 그릴 수 있는 기능까지 추가되고 있다.

시선 맞춤(eye contact) 기술은 영상회의에서 가장 중요한 기술 중 하나이다. 기존의 시스템은 화자의 시선과 카메라의 렌즈의 위치가 달라 화자 사이의 시선

* 광주과학기술원 정보통신공학과 (sblee, hoyo@gist.ac.kr) (° : 교신저자)

논문번호 : KICS2012-02-059, 접수일자 : 2012년 2월 11일, 최종논문접수일자 : 2012년 4월 10일

불일치가 발생한다. 시선 불일치 문제는 화자끼리의 대화의 집중력 및 몰입감을 떨어뜨리기 때문에 이러한 문제를 해결하기 위해 많은 연구 기관들에 의해 시선 맞춤을 위한 연구가 진행되었다.

최근, 독일의 HHI 연구소에서는 앞서 언급한 주요 기술들을 이용한 3차원 원격 영상회의 시스템을 개발했다^{5,6}. 이 방법은 HHI는 3대의 L 프레임 카메라와 디스플레이 하단의 1대의 카메라로부터 얻은 영상으로부터 스테레오 정합 및 visual hull 알고리즘을 동시에 사용하여 영상을 합성하는 시선 조정 기술을 개발하였다. 하지만, 기존의 방법들은 성능에 비해 하드웨어 구성이 너무 복잡하고 시스템 구축에 너무 많은 비용이 든다는 단점이 있다.

본 논문에서는 깊이 영상의 후처리 필터링 기술을 통한 원격 영상회의의 시선 맞춤 방법을 제안한다. 제안하는 방법은 기존의 깊이탐색 기술에서 발생하는 객체의 경계 불일치, 시간적 상관도 저하 등의 문제들을 해결하기 위해, 깊이 영상에 대해 색상 영상의 경계 정보와 색상 정보를 깊이 영상의 필터링에 사용하는 결합형 양방향 필터(joint bilateral filter)를 사용한다. 제안하는 방법을 통해 카메라와 화자와의 거리를 계산할 수 있게 되고, 최종적으로 정면시점 영상합성 방법을 통해 화자가 정면을 바라보는 듯한 영상을 합성해서 시선 맞춤을 가능하게 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안하는 시선 맞춤 시스템의 개요를 소개하고, 3장에서는 깊이 영상의 화질을 향상시키기 위한 깊이 영상의 후처리 필터링 기술을 설명한다. 4장에서 실험 결과를 통해 제안하는 시스템의 성능을 분석한 다음 5장에서 결론을 맺는다.

II. 시스템 개요

그림 2. 시스템 전체 블록 다이어그램
Fig. 2. Block diagram of the proposed system

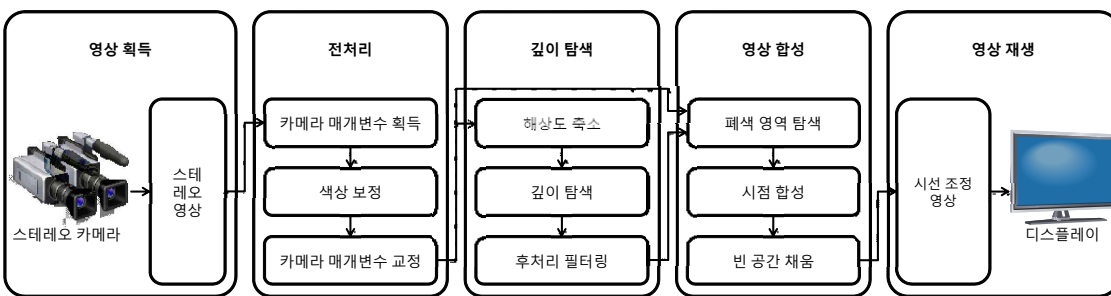


그림 1은 제안하는 알고리즘을 위한 시스템 구성 환경을 나타낸다. 그림 1에서 알 수 있듯이, 제안하는 기술은 두 대의 카메라를 디스플레이의 좌우 각각에 배치시키고 카메라의 광축이 화자로 향하는 수렴형 카메라 배열을 사용한다. 또한, 카메라의 위치는 디스플레이 상에서 화자의 시선이 위치하는 높이에 맞도록 설치된다.

수렴형 배열의 스테레오 카메라

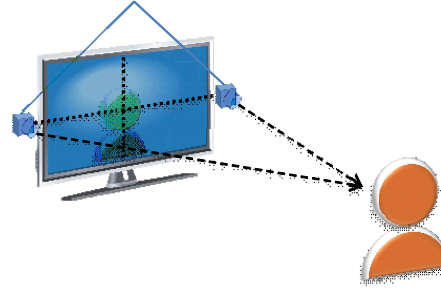


그림 1. 시스템 구성 환경
Fig. 1. System configuration

그림 2는 제안하는 시스템의 전체 블록 다이어그램을 나타낸다. 제안하는 시스템의 기술 흐름은 다음과 같이 요약할 수 있다. 첫번째로 수렴형 배열로 구성된 멀티 카메라를 이용해서 화자를 촬영한다. 두 번째 단계로 스테레오 영상의 전처리 과정을 거치는데, 격자 무늬의 패턴이 촬영된 여러 장의 2차원 영상들에서 추출한 특징점을 기반으로 카메라의 내부 매개변수와 외부 매개변수를 계산하는 카메라 보정 (camera calibration) 기술, 스테레오 영상 사이의 색상 차이를 제거해주는 색상 보정 기술, 스테레오 영상의 기하학적 오차를 최소화하는 카메라매개변수 교정

(rectification) 기술을 사용한다. 깊이 영상을 탐색하는 세 번째 단계에서는 계산량을 감소시키기 위해 스테레오 영상의 해상도를 축소하고 깊이탐색 방법을 통해 카메라와 화자와의 거리를 계산한 다음 후처리 필터링을 거친다. 시선이 조정된 영상을 합성하는 네 번째 단계에서는 합성될 시점에서의 폐색 (occlusion) 영역을 탐색하고, 가상 시점 영상을 합성한 다음, 채워지지 않은 빈 공간들을 주변 화소들을 이용해서 채워서 시선이 조정된 영상을 얻는다⁷⁾. 최종 단계에서 디스플레이를 통해 상대방에게 시선이 조정된 영상을 재생하여 자연스러운 시선 맞춤을 구현한다.

III. 제안하는 깊이 영상의 후처리 필터링 기술

깊이탐색 기술을 이용하면 3차원 장면의 깊이 정보를 획득할 수 있지만, 그림 3에서 알 수 있듯이 색상 영상과 변위 영상의 객체 경계가 맞지 않는 문제점이 발생한다. 이 문제점은 전역적 방법을 사용했을 때 자주 발생하는 것으로 영상을 합성하는 과정에서 오차를 발생시킬 뿐만 아니라, 다양한 3차원 응용 분야에 깊이 정보를 사용할 수 없게 한다.

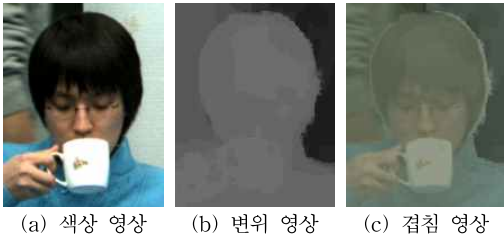


그림 3. 색상 영상과 변위 영상의 경계 불일치
Fig. 3. Boundary mismatch between color image and depth map

최근, 이와 같은 문제점을 해결하기 위해서 획득한 변위 영상에 대해 후처리 필터링을 적용하는 방법들이 제안됐다^{8,9)}. 후처리 필터링은 기본적으로 양방향 필터(bilateral filter)로부터 출발한다. 기존의 양방향 필터는 한 장의 색상 영상에 적용되는 지역 필터로써, 인접 화소와의 거리차, 색상차 각각에 대한 두 개의 가우시안 (Gaussian) 함수를 사용한다. 새롭게 제안된 결합형 양방향 필터(joint bilateral filter)에서는 두 개의 가우시안 함수에 추가적으로 인접 화소와의 변위 차에 대한 항을 사용한다. 그림 4는 변위 영상에 후처리 필터링을 적용한 결과를 나타낸다. 그림 4(a)와 그림 4(b)에서 알 수 있듯이, 필터링을 적용하기 전 변위 영상에 비해 객체의 경계가 상당 부분 일치하는 것을 확인할 수 있다.



(a) 후처리 전 영상 (b) 후처리 후 영상
그림 4. 결합형 양방향 필터링 결과
Fig. 4. Result for joint bilateral filtering

본 논문에서는 기존 방법에서 사용한 양방향 필터를 시간축으로 확장하는 깊이 영상 후처리 필터링 방법을 제시한다. 제안하는 방법은 필터 적용 범위를 시간축으로 확장한 3차원 필터를 사용한다. 확장된 결합형 양방향 필터는 다음과 같이 정의된다.

$$D(x, y, t) = \arg \min_{d \in \hat{d}_p} \frac{\sum_{u \in \Gamma_p} \sum_{v \in \Gamma_p} \sum_{w \in \Gamma_p} W(u, v, w) C(u, v, w, d)}{\sum_{u \in \Gamma_p} \sum_{v \in \Gamma_p} \sum_{w \in \Gamma_p} W(u, v, w)} \quad (1)$$

위 식에서 각 변수는 $p=(x, y)$, $\hat{d}_p=\{D(x-1, y, t), D(x+1, y, t), D(x, y-1, t), D(x, y+1, t), D(x, y, t-1), D(x, y, t+1)\}$, $u_p=\{x-r, \dots, x+r\}$, $v_p=\{y-r, \dots, y+r\}$, $w_p=\{t-r, \dots, t+r\}$ 와 같고 $D(x, y, t)$ 는 변위값을 나타낸다. 위 식에서 양방향 필터의 가중값인 $W(u, v, w)$ 와 정합 비용 $C(u, v, w, d)$ 는 다음과 같이 정의된다.

$$W(u, v, w) = \exp\left\{-\frac{\|I(x, y, t) - I(u, v, w)\|^2}{2\sigma_r^2}\right\} \times \exp\left\{-\frac{\sqrt{(x-u)^2 + (y-v)^2 + (t-w)^2}}{2r^2}\right\} \quad (2)$$

$$C(u, v, w, d) = \min(\lambda L, |D(u, v, w) - d|) \quad (3)$$

여기서 λ 는 정합 오차 $C(u, v, w, d)$ 의 상한을 결정하기 위한 상수를 나타내고 Γ 은 스테레오 정합에 사용된 변위값의 범위를 나타낸다. 그리고 $I(x, y, t)$ 는 색상 영상의 화소값을 나타낸다.

위와 같은 결합형 양방향 필터를 확장해서 깊이 영상에 적용하게 되면 깊이 영상 내의 대부분의 영역의 시간적 상관도를 향상시킬 수 있지만 움직임이 급격한 객체에 대해서는 필터링 이후에 오히려 깊이값의 오차가 발생할 수 있다. 그림 5는 시간축으로 확장만 했을 때의 필터링 이후의 깊이값 오차를 나타낸다.

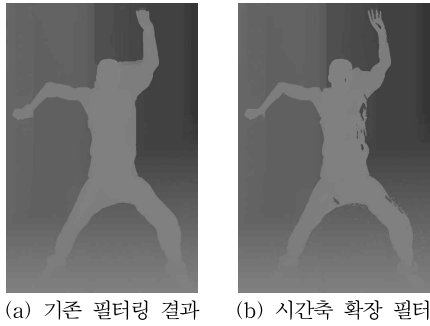


그림 5. 움직임이 급격한 객체의 깊이값 오차
Fig. 5. Depth error for dynamic object

따라서, 본 논문에서는 이를 해결하기 위해 시간축으로 발생하는 외곽 오차(outlier)를 제거하기 위한 방법을 추가적으로 적용한다. 제안하는 방법은 양방향 필터의 3차원 윈도우 내의 이전 화면과 이후 화면에서 (x,y) 위치의 깊이값을 조사한다. 이 깊이값들 가운데 현재 화면과의 차이가 한계값 이상이고 화소값 차이 역시 또 다른 한계값 이상일 경우에는 외곽 오차가 포함된 화면이라 판단하여 이 화면을 제거한 후 필터링을 수행한다. 이를 수식으로 나타내면 다음과 같다.

$$w_{outlier_reduction} = \left\{ w_p \mid \begin{aligned} &|I(x,y,t) - I(x,y,w_p)| < 2\lambda L \\ &|D(x,y,t) - D(x,y,w_p)| < \lambda L \end{aligned} \right\} \quad (4)$$

IV. 실험 결과 및 분석

제안하는 시스템은 스테레오 카메라만을 이용하기 때문에 원본 깊이 영상을 획득할 수 없다. 그렇기 때문에, 제안하는 방법의 성능을 평가하기 위한 깊이 영상의 화질 비교는 원본 깊이 영상이 함께 제공되는 테스트 영상을 통해서만 가능하다. 제안한 깊이 영상 필터링 방법의 성능을 평가하기 위해, MPEG의 3차원 비디오 부호화 그룹의 테스트 영상 가운데 하나인 "Undo_Dancer" 영상을 사용했다. 이 영상은 해상도가 1920x1080이며 컴퓨터 그래픽 작업을 통해 얻은 영상으로 원본 깊이 영상이 함께 제공됐기 때문에, 깊이 영상의 화질을 객관적으로 비교할 수 있다. 제안한 방법의 성능은 원본 깊이 영상, 깊이탐색 알고리즘을 통해 얻어진 깊이 영상, 기존의 필터링 방법을 적용한 결과, 제안한 필터링 방법을 적용한 결과의 화질을 비교하는 방법을 사용했다.

그림 6은 다양한 방법을 통해 획득한 깊이 영상을 나타낸다. 그림 6(b)는 색상 영상을 이용해서 그래프 컷 기반의 깊이탐색 방법을 적용했을 때 결과를 나타

낸다¹⁰⁾. 그래프 컷 방법을 이용해서 얻은 깊이 영상에 기존의 필터링 방법과 제안하는 방법을 적용했을 때 그림 6(c)와 그림 6(d)에서 볼 수 있듯이, 제안하는 방법이 객체의 경계를 보다 명확하게 일치시켜 주는 것을 알 수 있었다.

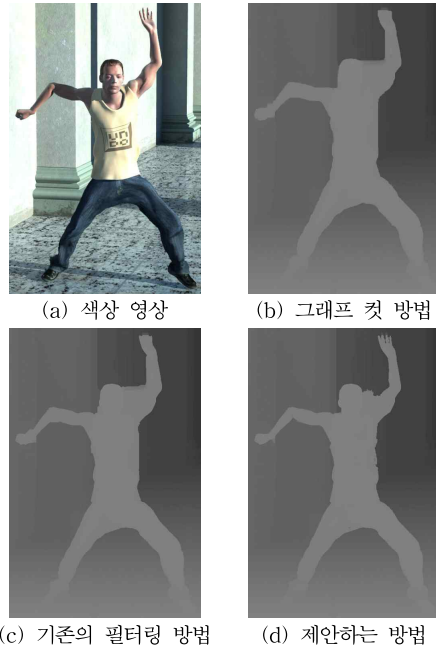


그림 6. 깊이 영상 후처리 필터링 결과
Fig. 6. Results for depth map filtering

그림 7은 제안하는 방법의 성능을 객관적으로 평가하기 위해, 원본 깊이 영상과 매 화면마다 화질 비교를 수행했다. 화질은 PSNR값으로 비교했다. 그림 7에서 알 수 있듯이, 제안하는 방법이 기존의 필터링 방법에 비해 보다 높은 PSNR로 깊이 영상을 필터링했음을 알 수 있었다. 표 1은 화질의 평균값을 보여준다. 제안한 방법을 사용했을 때, 기존의 방법에 비해 약 2.33 dB의 화질을 향상시켰다.

표 1. 깊이 영상 화질의 평균값 비교
Table 1. Average PSNR of depth quality

| 방법 | 그래프 컷 방법 | 기존 필터링 방법 | 제안하는 방법 |
|-----------|----------|-----------|---------|
| PSNR (dB) | 37.5499 | 37.9307 | 40.2620 |

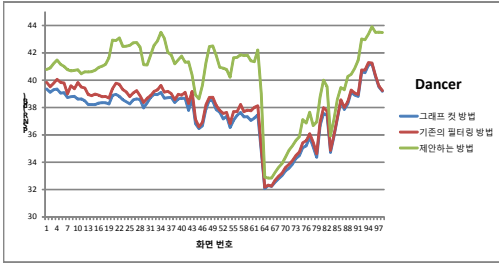


그림 7. 깊이 영상 화질 비교
Fig. 7. Comparison of depth quality

본 논문에서 제안하는 시선 맞춤 방법의 성능을 평가하기 위해, 그림 8과 같은 시스템을 구축하고 영상을 획득했다. 카메라의 간격은 42인치 디스플레이를 기준으로 했을 때, 약 93cm, 시청거리는 디스플레이의 인치별 권고 시청거리에 맞도록 2.7m로 설정했다^[11]. 영상을 획득하기 위해서 배경의 변화는 없고 배경과 화자의 거리는 거의 차이가 없도록 제한사항을 두었다. 카메라 모델은 Point Grey Research 사의 Grasshopper이며 해상도는 1280×960으로 설정했다.



그림 8. 시선 맞춤을 위한 스테레오 카메라 시스템
Fig. 8. Stereo camera system for eye gaze correction

그림 9는 깊이 영상의 후처리 필터링 적용 유무에 따른 정면시점 합성영상의 화질을 보여준다. 그림 9에서 좌측 2개 영상은 좌시점 카메라, 우시점 카메라 각각에서 획득된 영상이며, 우측 영상은 정면시점으로 영상을 합성한 결과이다. 그림 9(a)와 비교해보면, 그림 9(b)에서 알 수 있듯이 제안하는 깊이 영상 필터링 기술을 사용했을 때, 화자의 경계 부분이 더 자연스러운 것을 확인할 수 있었다. 화질 비교를 명확하게 하기 위해서 그림 10에서는 화자의 경계 부분을 확대한 영상을 보여준다.



(a) 필터링 적용 전



(b) 필터링 적용 후

그림 9. 시선이 조정된 합성영상
Fig. 9. Gaze-corrected virtual image



(a) 필터링 적용 전 (b) 필터링 적용 후

그림 10. 시선 조정 영상의 확대 영상
Fig. 10. Zoom-in images of gaze-corrected image

V. 결 론

본 논문에서는 원격 영상회의의 시스템의 시선 맞춤을 위한 깊이 영상 필터링 기술을 제안했다. 제안한 시스템은 기존의 방법에 비해 하드웨어 구성의 복잡도가 단순하면서도 시선 맞춤 문제를 해결할 수 있었다. 또한, 깊이 영상에 대해 시간축으로 확장한 결합형 양방향 필터를 사용함으로써, 객체의 경계 불일치, 시간적 상관도 저하 등의 문제 등을 해결할 수 있었다. 제안한 후처리 기술을 통해 정제된 깊이영상을 이용해서 정면시점에서의 영상을 합성했기 때문에, 기존의 방법에 비해 약 2.33 dB의 화질 향상 효과를 볼 수 있었고, 보다 자연스러운 합성영상을 얻을 수 있었다. 결과적으로, 화자 사이에 시선 맞춤이 가능한 것을 확인할 수 있었다.

참 고 문 헌

- [1] P. Kauff and O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," Proc. of international conference on Collaborative virtual environments, pp. 105-112, Sep. 2002.
- [2] S. Huang and J. Wang, "A low-cost desktop videoconferencing codec: an adaptive Motion-JPEG design," IEEE Transactions on Consumer Electronics, vol. 40, no. 4, pp. 944-950, Nov. 1994.
- [3] S. M. Kuo, Y. C. Huang, and Z. Pan, "Acoustic noise and echo cancellation microphone system for videoconferencing," IEEE Transactions on Consumer Electronics, vol. 41, no. 4, pp. 1150-1158, Nov. 1995.
- [4] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, "New video enhancement preprocessor using the region-of-interest for the videoconferencing," IEEE Transactions on Consumer Electronics, vol. 56, no. 4, pp. 2644-2651, Nov. 2010.
- [5] O. Schreer, N. Atzapadin, and I. Feldmann, "Multi-baseline disparity fusion for immersive videoconferencing," In Proceedings of International Conference on Immersive Telecommunications, pp. 27-29, May 2009.
- [6] P. Kauff and O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," In Proceedings of International Conference on Collaborative Virtual Environments, pp. 105-112, Oct. 2002.
- [7] S. Lee, I. Shin, and Y. Ho, "Gaze-corrected View Generation using Stereo Camera System for Immersive Videoconferencing," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1033-1040, 2011.
- [8] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.
- [9] P. Lai, D. Tian, and P. Lopez, "Depth map processing with iterative joint multilateral filtering," in Proceedings of Picture Coding Symposium, pp. 9-12, 2010.
- [10] ISO/IEC JTC1/SC29/WG11, "Reference Software of Depth Estimation and View Synthesis for FTV/3DV," M15836, Oct. 2008.
- [11] M. Sugawara, K. Mitani, M. Kanazawa, F. Okano, and Y. Nishida, "Future prospects of HDTV - technical trends toward 1080p," SMPTE Journal, vol. 115, pp. 10-15, 2006.

이 상 범 (Sang-Beom Lee)

정회원



2004년 경북대학교 전자전기공학부 졸업(학사)
 2006년 광주과학기술원 정보통신공학과 졸업(석사)
 2006년~현재 광주과학기술원 정보통신공학과 박사과정 <관심분야> 3차원 TV, 실감방송, 3차원 비디오 부호화

호 요 성 (Yo-Sung Ho)

정회원



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 정보통신공학과 교수
 <관심분야> 디지털 신호처리, 영상 신호 처리 및 압축, 디지털 TV와 고선명 TV, 멀티미디어 시스템, MPEG 표준, 3차원 TV, 실감방송