

2D/3D 변환을 위한 깊이정보 생성기법에 관한 연구

한현호¹, 이강성², 이상훈^{2*}
¹광운대학교 대학원, ²광운대학교 교양학부

A Study on 2D/3D image Conversion Method using Create Depth Map

Hyeon-Ho Han¹, Gang-Seong Lee² and Sang-Hun Lee^{2*}

¹Kwangwoon University Graduate School

²Kwangwoon University General Education

요 약 본 논문은 2D/3D 변환을 위한 객체 추출과 깊이정보(Depth-map) 생성기법에 관한 연구이다. 2D영상을 3D로 변환하기 위해서는 영상 객체 추출, 영상 거리 인식, 영상 생성, 재보정 단계를 거치게 되는데 본 논문에서는 영상 객체 추출과 영상 거리 인식에 해당하는 깊이정보를 생성하는 방법을 제안한다. 3D 영상으로의 변환은 2D 영상에서의 객체 추출과, 추출된 객체와 주변 배경을 구별하기 위한 거리감을 할당하는 깊이정보 생성이 중요하다. 보다 정확한 객체 추출과 깊이정보를 생성하기 위해 기존의 Optical flow에서 잡음을 제거한 방법을 제안하였다. 제안한 방법으로 2D 영상을 깊이정보가 포함된 영상으로 변환하여 영상의 깊이 정보가 추정됨을 알 수 있다.

Abstract This paper discusses a 2D/3D conversion of images using technologies like object extraction and depth-map creation. The general procedure for converting 2D images into a 3D image is extracting objects from 2D image, recognizing the distance of each points, generating the 3D image and correcting the image to generate with less noise. This paper proposes modified new methods creating a depth-map from 2D image and recognizing the distance of objects in it. Depth-map information which determines the distance of objects is the key data creating a 3D image from 2D images. To get more accurate depth-map data, noise filtering is applied to the optical flow. With the proposed method, better depth-map information is calculated and better 3D image is constructed.

Key Words : 2D/3D Conversion, Optical Flow, Object Segmentation, Depth map

1. 서론

과학자들의 연구에 의해 인간은 양안의 시차 즉, 물체가 좌, 우안에 투영될 때 생기는 영상에 차이가 있으며 이 좌, 우 두 영상이 뇌에서 합성되어 3차원 입체 영상으로 인식하게 된다는 것을 알게 되었다. 그 결과로 3차원 입체 영상을 보기 위해 좌, 우 영상을 생성한 뒤 각각 좌, 우안에 보여주는 방법이 고안되었고 이 방법을 통해 다양한 3차원 입체 영상을 생성하게 되었다[1].

최근에는 3차원 디스플레이의 발전과 보급화가 이루

어지면서 그로 인해 다양한 3차원 콘텐츠의 생산이 요구되고 있다.

일반적으로 입체 영상 기술은 2대의 카메라로 구성된 스테레오 카메라로 촬영하거나 이미지 편집 툴을 이용하여 획득하였다. 하지만 스테레오 카메라를 이용한 제작은 많은 비용이 필요하기 때문에 한계가 있고, 이미지 편집 툴을 이용할 경우 많은 시간이 소요된다. 이러한 일반적인 입체 영상 획득 기술에 반해 2차원 영상의 3차원 입체 변환 기술은 1대의 카메라로 촬영한 영상을 입체 변환 기술을 이용하여 자동적으로 입체 영상을 생성하는 기술이

본 논문은 2009년도 광운대학교 교내 학술 연구비 지원에 의해 연구 되었음

*교신저자 : 이상훈(leesh58@kw.ac.kr)

접수일 11년 03월 04일

수정일 (1차 11년 03월 15일, 2차 11년 04월 06일)

게재확정일 11년 04월 07일

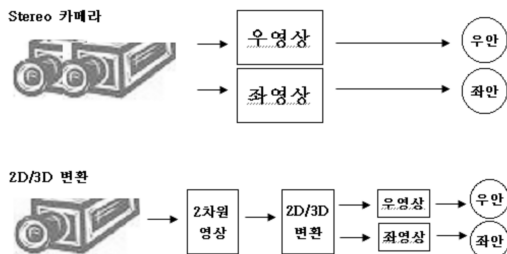
다. 단일 카메라를 이용해 제작된 많은 영상들을 변환하여 3차원 디스플레이에서 활용할 수 있다면 3차원 디스플레이가 상용화 된다 해도 사용자의 요구를 충분히 만족시킬 수 있을 것이다. 따라서 단일 카메라로 제작된 영상을 입체 영상으로 변환하는 기법에 대해 많은 연구가 이루어지고 있다[2].

2D/3D 변환 기법은 단일 카메라로 제작된 영상에 Ross 현상을 기반으로 연속된 프레임 중에 좌, 우안에 각각 비춰줬을 때 입체로 느낄 만큼의 시간지연을 갖는 2개의 프레임을 합성하여 보여주는 방법이 대표적이다[3]. 이를 응용하여 객체와 영역을 분할하고 추출한 뒤, 지연 요소를 연산하여 선별된 프레임들을 합성하는 MTD (Modified Time Difference) 방법과, 인간의 시각적 특성인 공간 시간 보간을 이용한 방법, 운동시차를 이용하여 영상의 깊이 정보를 생성하는 방법 등이 있다[4,5]. 하지만 기존의 방법들은 이미지 변형에 의해 영상의 화질이 떨어지고, 정확성이 떨어지는 단점이 있다.

본 논문에서는 객체 분할 정보를 획득, 추출하고 잡음을 제거한 Optical Flow를 이용하여 얻은 움직임 정보를 통해 깊이정보를 생성하고 각 객체에 깊이 정보를 할당하는 방법을 제안하였다.

2. 2D/3D 영상 변환 기법

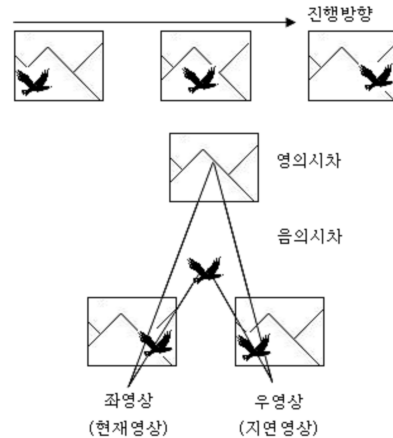
2D/3D 영상 변환 기술은 기존의 단일 카메라로 촬영된 영상 즉, TV나 CD, DVD에서 얻어지는 영상을 영상 처리 기술을 통해 좌영상과 우영상으로 구성된 입체 영상을 생성하는 것으로 2대의 카메라로 이루어진 스테레오 카메라를 이용하여 얻어진 입체 영상을 디스플레이 기기에 보내주는 입체 영상 처리와는 달리 상대적으로 복잡한 원리 및 영상 처리 기술을 필요로 한다. 2D/3D 영상 변환에서 좌영상과 우영상의 생성은 현재 영상과 이전 영상을 이용하여 생성하는 것이 다수의 시스템에서 사용하고 있는 방식이다[6].



[그림 1] 2D/3D 영상 변환의 원리

2.1 Ross 현상

Ross 현상을 이용한 영상 변환은 가장 기본적인 입체 변환 기법으로서, 사람은 양 눈의 시각적 차이로 인해 입체감을 느낄 수 있다는 이론을 기반으로 2차원 영상에서 시간적인 지연을 갖는 좌, 우 영상을 선택하여 각각 좌, 우안에 비춰주게 되면 움직이는 객체는 스크린보다 앞쪽에 양안이 수렴되어 입체감을 느낄 수 있다.



[그림 2] Ross 현상

그림 2는 Ross현상을 설명하는 예이다. 영상은 고정된 배경의 산과 좌에서 우로 이동하는 새이다. 이 경우 좌영상에 현재영상을 사용하고 우영상은 지연영상을 사용하면 배경이 되는 산은 스크린에 수렴점이 생기고 새는 스크린보다 앞쪽에 수렴점이 생성되어 입체감을 느끼게 된다.

표 1은 Ross 현상을 적용하기 위해 영상에서의 다양한 물체 및 카메라 운동유형에 따른 지연영상과 현재영상의 선택법을 나타내고 있다.

[표 1] 운동유형에 따른 영상 선택 방법

운동 종류	카메라 운동	객체 운동	좌영상	우영상
①	우측	없음	지연영상	현재영상
②	좌측	없음	현재영상	지연영상
③	없음	우측	현재영상	지연영상
④	없음	좌측	지연영상	현재영상
⑤	우측	우측	현재영상	지연영상
⑥	우측	좌측	지연영상	현재영상
⑦	좌측	좌측	지연영상	현재영상
⑧	좌측	우측	현재영상	지연영상

만약 좌영상과 우영상이 잘못 선택된 경우에는 왜곡된 시차를 얻게 되어 스크린보다 뒤쪽에 수렴점이 생기게 되고 배경보다 들어가 보이게 되므로 왜곡된 입체감을 느끼게 한다. 따라서 카메라와 객체의 운동유형에 따른 좌우영상 선택 방법이 필요하며 각 객체의 운동유형을 결정하는 기법이 필요하게 된다.

또한 Ross현상은 두 영상간의 수직시차가 존재하는 경우에 시간적으로 지연된 영상을 선택해서 비춰주는 방법을 이용하기 때문에 수렴점이 존재하지 않으므로 수직시차가 존재하는 영상에는 적용하기 어려운 단점이 있다.

2.2 MTD (Modified Time difference)

MTD는 객체를 추출하고, 움직임이 있는 영역을 분리한 후, 그 영역을 추출하여 현재 영상과 지연된 영상의 운동방향, 유형을 판단하여 지연 요소를 구하고, 지연 요소 값에 의해 합성할 좌, 우 영상을 결정한 뒤, 각각 좌, 우안에 비춰주어 입체로 느끼게 하는 방법이다. MTD 방법은 운동방향이 수평인 경우 외에도 운동유형과 운동방향을 고려한 변환을 통해 좌, 우 영상을 생성이 가능하다.



[그림 3] MTD 기법을 이용한 변환 과정

하지만 MTD방법은 이미지 변형에 인해 화질이 떨어지고 블록매칭을 통해서 운동정보를 얻으므로 정확성이 떨어지는 단점이 있다.

이러한 단점을 보완하기 위해 기존 Optical Flow에서 잡음을 제거한 방법을 이용하여 객체 분할 알고리즘인 Normalized Cut에 가중치로 적용하고 분할된 객체에 객체별 깊이 정보를 할당하는 방법을 제안하였다.

3. 객체 추출

본 논문에서 제안하는 2D/3D 변환 기법을 적용하기

위해 먼저 객체 추출 과정을 거치게 된다. 객체 추출은 그래프 분할 알고리즘인 Normalized cut을 이용한다[7]. Normalized cut은 영상을 그래프로 표현한 뒤 그래프 분할 과정을 거치고 분할된 결과를 영상에 대응시키는 알고리즘이다. 이를 적용하기 전에 연산량 감소를 위한 Watershed 알고리즘을 적용하고 객체 분할의 정확도 향상을 위해 잡음을 제거한 Optical flow를 추가 적용함으로써 보완하였다.

3.1 초기 영역 분할

객체 분할 과정의 연산량 감소를 위해 초기 영역 분할으로서 먼저 영상의 에지 추출을 통해 Gradient map을 생성하게 된다. Gradient map을 생성하기 위해 추출한 원 영상을 Gray scale로 변환하여 각 픽셀이 0~255의 색상값을 갖도록 한 후 Sobel 마스크를 이용하여 영상의 에지를 추출하였다.



[그림 4] 원본 이미지와 Gradient map 이미지

얻어진 Gradient map을 이용하여 밝기 정보의 변화가 큰 곳에서 영상을 분할해내는 Watershed 알고리즘을 적용하여 영상의 초기 영역 분할 과정을 거치고 분할 영역에 각각 평균 밝기값을 할당하고 영역단위로 연산하게 되면 기존 픽셀단위 연산에 비해 연산량이 감소하게 된다[8].



[그림 5] 초기 영역 분할과 평균 밝기 할당

3.2 Normalized cut을 이용한 객체 분할

Normalized cut은 전체 영상의 픽셀을 노드 u, v 로 정

의하고 픽셀 간의 유사도 $w(u,v)$ 를 에지로 가지는 그래프 $H = \{(i, j), w(u,v)\}$ 로 표현한다. 그래프 분할은 노드 u, v 간의 가중치 값 $w(u, v)$ 의 합인 $cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$ 을 통해 이루어진다.

Normalized cut은 식 (1)과 같이 특정 집합 A 에 속하는 점에서 그래프에 존재하는 모든 점의 집합 V 까지의 가중치 합 $assoc(A, V)$ 를 구한다.

$$assoc(A, V) = \sum_{u \in A, t \in V} w(u, t) \quad (1)$$

특정 집합 B 에 적용될 $assoc(B, V)$ 도 동일하게 정의하면 Normalized cut의 전체 수식은 식 (2)과 같다.

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \quad (2)$$

이 방법은 영상을 분할할 때 기존의 단순하게 분할하던 방법인 $cut(A, B)$ 를 보완하여 집합 A, B 에 대해 각각 $assoc(A, V), assoc(B, V)$ 로 나누어서 정규화 과정을 거치므로, 기존의 $cut(A, B)$ 이 따로 떨어진 노드가 작은 값을 가지게 되어 따로 분할되었던 것과 달리 $Ncut(A, B)$ 값은 따로 떨어진 노드를 보다 큰 값을 가지게 하여 잘못된 분할을 방지한다. 따라서 $Ncut(A, B)$ 가 가장 작은 값을 가지게 되는 A, B 를 찾아 분할하여 영상 내에서의 객체를 분할해 낼 수 있다.

하지만 Normalized cut이 정확한 분할을 하기위해 가중치의 선정이 중요하다. 식 (3)은 가중치 값인 밝기 차이와 인접 정도를 나타내기 위한 식으로, $F(u)$ 는 u 번째 픽셀에서의 밝기이고, σ_U 는 전체 영상 밝기의 표준편차이며 σ_X 는 전체 영상 위치의 표준편차, $X(u)$ 는 u 번째 픽셀의 위치, r 은 인접한 범위를 나타낸다.

$$w_{uv} = \exp \left(\frac{-|F(u) - F(v)|_2^2}{\sigma_U^2} \right) * \begin{cases} \exp \left(\frac{-|X(u) - X(v)|_2^2}{\sigma_X^2} \right), & \text{if } \|X(u) - X(v)\|_2 < r \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

하지만 픽셀 간 밝기 차이와 인접 정도만으로는 복잡한 영상에 적용할 시 원활한 결과를 보여주지 못하므로 추가적인 가중치가 필요하다.

4. 제안하는 깊이정보 생성 방법

4.1 잡음을 제거한 Optical flow

Optical flow는 연속적인 영상에서 동일한 객체의 밝기 값이 일정하게 유지되는 것을 이용하여 객체의 움직임을 찾아내는 알고리즘이다[9]. Optical flow는 픽셀단위로 움직임 정보를 추정함으로써 블록 매칭 같은 알고리즘에 비해 정확도가 높다. 본 논문에서는 이러한 장점을 이용하여 보다 정확한 객체 분할을 위해 가중치로서 이용한다.

다음 식 (4)은 Optical flow의 전체 수식으로, $o(u, v)$ 는 폐색 변수이며 $I(u, v)$ 는 LAB 색공간의 벡터를 나타낸다[10].

$$Oflow = \exp \left(- \frac{|u - u'|^2 + |v - v'|^2}{2\sigma_1^2} - \frac{|I(u, v) - I(u', v')|^2}{2\sigma_2^2} \right) \frac{o(u', v')}{o(u, v)} \quad (4)$$



[그림 6] 원본이미지와 Optical Flow 결과 이미지

하지만 Optical flow는 촬영된 영상에서 카메라의 이동이나 전체 조명의 변화가 지속되는 경우 그로인해 불필요한 정보까지 함께 연산되어 정확한 움직임 정보를 얻기가 힘들다.

다음 식 (5)는 이러한 점을 보완하기 위해 기존 Optical flow에서 잡음을 제거하는 수식으로, σ_Y 는 전체 영상의 Optical flow 표준편차, $Y(u)$ 는 u 번째 픽셀의 Optical flow 값이며 y 는 추정 Optical flow값의 오차범위이다. 이 값을 통해 움직임 정보가 필요한 부분 외에 불필요한 정보를 제거하게 된다.

$$FOflow = \begin{cases} \exp \left(\frac{-|Y(u) - Y(v)|_2^2}{\sigma_Y^2} \right), & \text{if } \|Y(u) - Y(v)\|_2 < y \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

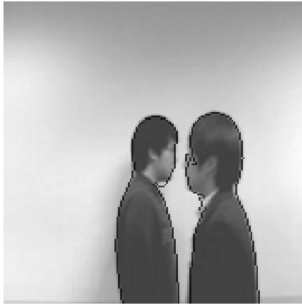
결과적으로 제안하는 기법의 최종 변환 과정은 다음 그림 10과 같다.



[그림 7] 잡음이 제거된 Optical flow 결과

얻어진 결과를 Normalized cut의 가중치로 적용시키면 다음 그림 8과 같은 결과를 얻을 수 있다.

$$OFNcut(A, B) = \frac{\sum_{u \in A, v \in B} w(u, v)}{\sum_{u \in A, t \in V} w(u, t)} + \frac{\sum_{u \in A, v \in B} w(u, v)}{\sum_{v \in B, t \in V} w(v, t)} \quad (6)$$



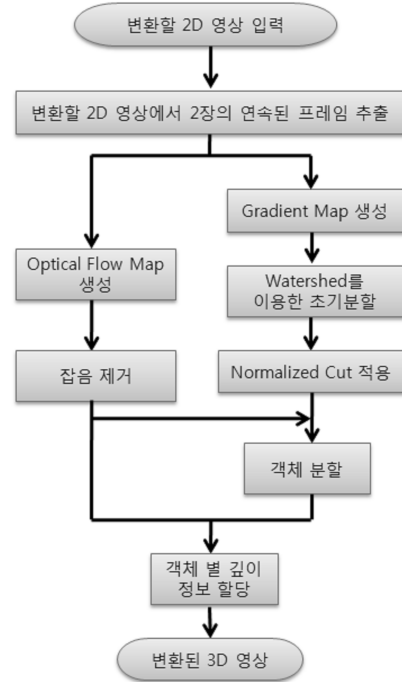
[그림 8] 최종 객체 분할 결과 이미지

4.2 추출된 객체의 깊이 정보 할당

최종 객체 분할 결과를 얻은 뒤 얻어진 객체 분할 정보를 바탕으로 깊이정보를 할당한다. 각 객체의 깊이정보는 Optical flow를 통해 얻어진 값을 기반으로 각 객체에 할당함으로써 객체별 깊이정보를 생성하게 된다[11].



[그림 9] 최종 깊이정보 생성 결과

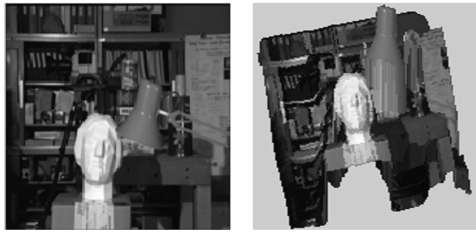


[그림 10] 제안하는 기법의 최종 변환 과정

5. 실험 및 고찰

제안한 방법을 검증하기 위해 실제 이미지에 적용하였다. 실험 환경은 MATLAB을 사용하여 시뮬레이션 하였다.





[그림 11] 2차원 영상의 3차원 변환 결과

Optical Flow를 추가한 Normalized Cut을 이용했을 때, 객체 분할 과정에서 약간의 잡음이 있지만 전체적인 객체 분할과 깊이 정보 추정이 가능한 것을 알 수 있다.

6. 결론

본 논문에서는 2차원 영상을 3차원 영상으로 변환하기 위해 잡음이 제거된 Optical Flow를 이용하는 방법을 제안하였다. 잡음을 제거한 Optical flow를 이용하여 Watershed 알고리즘을 선적용한 Normalized cut의 정확성을 향상시켰고, 추출된 객체에 Optical flow를 이용한 값을 할당함으로써 각 객체별 깊이 정보를 생성하였다.

제안한 방법은 Optical flow의 단점인 조명 변화와 카메라 이동에 대해 대처가 가능할 것으로 예상되며 디스플레이 방식에 따라 영상 생성 과정을 거친 뒤 사용할 수 있는 장점을 가지고 있기 때문에 기존 2차원 영상물을 기반으로 3차원 콘텐츠의 생성에 기여할 수 있을 것으로 예상된다.

향후 연구 방향으로는 디스플레이에 상관없이 사용할 수 있는 3차원 영상 생성과 자체적으로 재보정 과정을 포함하는 기법의 연구가 필요하다.

참고문헌

- [1] 이요섭, “2D-3D 변환 기술의 동향 및 전망”, 대한전자공학회, 전자공학회지, 제 38권, 2호, pp. 129-135, 2011.
- [2] B. J. Garcia, “Approaches to stereoscopic video based on spatiotemporal interpolation,”*Proc. SPIE*, vol. 2653, pp. 85-95, 1996.
- [3] T. Okino, H. Murata, K. Taima, T. Iinuma and K. Oketani, “New television with 2D/3D image conversion techniques,” *proc. SPIE*, vol. 2653, Stereoscopic Displays and Virtual Reality Systems III, 1996.
- [4] J. Ross, “tereopsis by binocular delay,”*Nature*, vol. 248, pp. 363-364, 1974.

- [5] Y. Matsumoto, H. Terasaki, K. Sugumoto, and T. Arakawa, “onversion system of monocular image sequence to stereousing motion parallax,”*Proc. SPIE*, vol. 3012, pp. 108-115, 1997.
- [6] 김만배, “2차원 동영상의 3차원 입체 변환 방법”, 한국방송공학회, 방송공학회지, 제6권 제1호 pp.26-35 2001.
- [7] Jianbo Shi and Jitendra Malik, “Normalized Cuts and Image Segmentation”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 22, NO. 8, 2000.
- [8] Luc Vincent and Pierre Soille, “Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations“, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 13, NO. 6, pp. 583-598, 1991.
- [9] Deqing Sun, Roth, S., Black, M. J., “Secret of Optical Flow Estimation and Their Principles”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2432-2439, 2010.
- [10] A. S. Ogale, C. Fermüller, and Y. Aloimonos, “Motion segmentation using occlusions,”*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 988-992, 2005.
- [11] 한현호, 홍영표, 김진수, 이상훈, “연산량을 감소한 객체 분할과 깊이정보 생성을 이용한 2D/3D 동영상 변환 연구”, 한국산학기술학회, 추계학술발표논문집, 제 11권, 제 2호, pp. 92-95, 2010.

한 현 호(Hyeon-Ho Han)

[준회원]



- 2009년 2월 : 광운대학교 컴퓨터 공학과 졸업(공학사)
- 2011년 2월 : 광운대학교 정보콘텐츠 대학원 유비쿼터스 컴퓨팅학과 졸업(공학석사)
- 2011년 ~ 현재 : 광운대학교 대학원 재학(박사과정)

<관심분야>

무선 네트워크, 3D영상처리

이 강 성(Gang-Seong Lee)

[정회원]



- 1986년 : 광운대학교 컴퓨터공학과 졸업(공학사)
- 1988년 : 광운대학교 대학원 컴퓨터공학과 졸업(공학석사)
- 1993년 : 광운대학교 대학원 컴퓨터공학과 졸업(공학박사)
- 2009년 : 서울대학교 음악대학원 작곡과 졸업(석사)

- CMU(Carnegie Mellon University), U.S.A. 교환교수
- 1990년 9월 ~ 현재 : 광운대학교 정교수

<관심분야>

음성공학, 음악공학, 미디어아트, 작곡

이 상 훈(Sang-Hun Lee)

[중신회원]



- 1983년 2월 : 광운대학교 응용전자 공학과 졸업(공학사)
- 1987년 8월 : 광운대학교 대학원 전자공학과 졸업(공학석사)
- 1992년 2월 : 광운대학교 대학원 전자공학과 졸업(공학박사)
- 1990년 9월 ~ 현재 : 광운대학교 정교수

- 2001년 ~ 2007년 : 세계기능경기대회(심사위원)
- 2006년 ~ 2007년 : 서울특별시 기능경기위원회(기술위원장)
- 2010년 ~ 현재 : 광운대학교 교양학부장

<관심분야>

무선인터넷, 무선네트워크, USN, 영상인식, 3D영상처리