

## 2차원 동영상의 3차원 동영상 변환

이 희 만\*

# Convert 2D Video Frames into 3D Video Frames

Hee man Lee \*

### 요 약

본 논문에서는 일반적인 2차원 비디오 프레임에 패러럴 스테레오 카메라로 촬영한 듯한 3차원 입체 비디오 프레임으로 변환하는 알고리즘을 제안하였다. 동영상의 연속하는 프레임 사이의 픽셀 변위를 구하고 이 변위 정보를 기반으로 입체영상을 합성한다. 픽셀 변위를 구하기 위하여 기존 수렴 반복법을 개선 보완하여 사용하였으며 픽셀변위 정보로부터 입체영상을 합성하는 방법을 제안하였다. 본 연구에서 제안한 알고리즘은 동영상의 어떤 유형이라도 일관성 있는 방법의 의하여 합성영상을 만들어 내므로 기존의 영상 유형 분류에 의한 합성 방법의 문제점을 개선하였다.

### Abstract

In this paper, An algorithm which converts 2D video frames into 3D video frames of parallel looking stereo camera is proposed. The proposed algorithm finds the disparity information between two consecutive video frames and generates 3D video frames from the obtained disparity maps. The disparity information is obtained from the modified iterative convergence algorithm. The method of generating 3D video frames from the disparity information is also proposed. The proposed algorithm uses coherence method which overcomes the video pattern based algorithms.

▶ Keyword : 입체영상변환(2D/3D Video Conversion), 변위맵(Disparity Map), 수렴 반복 알고리즘 (Iterative Convergence Algorithm)

---

• 제1저자 : 이희만

• 투고일 : 2009. 05. 29, 심사일 : 2009. 05. 29, 게재확정일 : 2009. 06. 21.

\* 서원대학교 멀티미디어학과 교수

## I. 서론

본 논문은 일반적인 2차원 비디오 프레임을 3차원 입체 비디오 프레임으로 구현하는 방법에 관한 것이다. 3차원 비디오 프레임은 기본적으로 2대의 비디오 카메라를 이용하여 얻어진다. 인간의 눈이 동일 사물을 2개의 눈으로 동시에 바라볼 때 각 망막에 비치는 물체 이미지의 위치 차이를 이용하여 입체를 인지하는데 3차원 동영상은 인간의 눈과 같이 2대의 카메라를 이용하여 촬영한 것으로 이 동영상을 인간의 좌우 눈에 각각 분리하여 보여줌으로서 입체영상을 인지 할 수 있도록 한다.

전자 전람회에서 가정용 3D 텔레비전이 대거 출시되고 있다. 이미 많은 3D 영화가 제작되고 있어 극장에서 상영 중이며 조만간 가정용으로 재배포될 전망이다. 우리나라에서는 세계 최초로 IPTV를 사용한 입체방송 시연에 성공하였다는 뉴스도 보도되고 있다. 3D 콘텐츠를 TV를 통해 3D 입체 콘텐츠를 즐길 수 있는 시대가 도래한 것이다. 사람의 양안시차를 이용한 3D 입체영상은 현재의 HD기술을 잇는 차세대 영상기술로 부각받고 있다. 그러나 3D입체영상을 제작하는 비용이 아직 고가이며 이용할 수 있는 콘텐츠도 매우 부족할 실정이다. 본 연구는 이미 다수 존재하는 일반 2D 비디오로부터 3차원 비디오 콘텐츠를 제작하는 방법에 관한 것이다.

## II. 선행연구

비디오는 많은 정보를 포함하고 있으므로 막대한 자료처리 및 정보의 이해에는 많은 어려움이 따른다[1]. 그러나 하드웨어의 비약적인 발전과 함께 비디오 정보의 처리에 대한 연구들이 활성화되고 있다. 본 연구는 이미 다수 존재하는 일반 2D 비디오를 정보처리하여 3차원 비디오 영상으로 변환한다.

일반 2D 비디오의 한 프레임은 한 대의 카메라를 이용하여 촬영한 것이다. 그러나 연속적인 비디오 프레임은 3차원 정보를 가지고 있다. 연속한 비디오 프레임은 비교적 짧은 시간 간격으로 촬영된 것으로 현재 프레임과 이전 프레임 간에는 사물과 카메라의 거리에 따라 이미지의 변화 정도가 각각 다르며, 또한 물체의 이동 속도에 따라 변화 정도가 달라진다. 즉 연속한 동영상을 여러 방법에 따라 각각 좌우 눈에 분리하여 보여주면 입체감을 느낄 수 있게 되는 것이다.

Chen은 동영상의 연속 프레임을 홀수 및 짝수번호로 분류하고 이를 좌측과 우측 눈에 각각 분리하여 보여주는 방법을 개발하였다[2]. 기존의 텔레비전의 NTSC 방식으로 입

체영상을 송출할 경우 초당 프레임 수가 1/2로 줄어들어 풀리커현상이 발생할수 있다. 이상용은 이런 문제를 해결하기 위해 기존 수직 주파수의 2배에 해당하는 동기신호를 발생하여 재변조하는 방식을 제안하였다[3]. 기존 프레임(예 좌측영상)은 정상적인 방식으로 보여주고 보조 프레임(예 우측영상)을 단순히 지연하는 방법도 제안되었다[4]. 또한 입체효과를 극대화하기 위해 비디오 프레임 상의 물체의 수평방향 이동이나 수직방향의 이동벡터의 정도에 따라 프레임의 지연 정도를 가변하는 방법도 제안되고 있다[5,6]. 프레임의 지연에 의한 입체 생성방식은 사물이 수직 방향으로 이동하거나 또는 카메라 쪽으로 다가오거나 멀어지는 장면에서는 효과적이지 못하다. 그러므로 동영상 장면의 이동벡터를 활용하여 보조 프레임을 합성하는 방식이 제안되었다[7,8]. 비디오 영상에서 이동벡터는 영상의 압축 뿐만 아니라 다양한 분야에서 이용되고 있다[9]. MPEG 동영상 스트림은 이동벡터를 사용하고 있으므로 이를 이용하여 변환하는 연구결과도 발표되었다. 김만배는 MPEG의 키프레임과 영상 블록 당 운동벡터(Motion Vector)로부터 카메라의 운동 유형을 파악하고 운동 유형에 따라 동영상 입체 변환을 한다[7]. Ianir Ideses는 MPEG 스트림으로부터 모션벡터 맵을 생성하고 이로부터 깊이 맵을 추정계산하고 리샘플링에 의한 동영상 입체변환을 제안하였다[8].

본 연구는 현재 프레임과 이전 프레임간 수렴 반복 알고리즘에 의해 변위 맵을 계산하고 리샘플링으로 보조 프레임을 생성하는 방식에 의해 2차원 동영상을 3차원 동영상으로 변환한다.

## III. 입체변환 원리

인간의 눈은 카메라와 매우 유사하다. 카메라는 실세계 3차원을 2차원의 필름에 상을 만드는 장치로 실세계에서의 사물의 위치 및 크기와 카메라의 필름에서 위치 및 크기와 관계는 투영변환으로 근사화 할 수 있다. 필름에 새겨진 상은 2차원으로 그 자체 만으로는 입체정보가 없다. 인간의 눈은 2개로 서로 일정 거리 분리되어 있으며 각 눈이 같은 사물을 각각의 망막에 투영할 수 있는데 각 눈의 위치가 서로 다르므로 같은 사물이 망막에 투영될 때 망막에서의 위치가 서로 다르게 된다. 이 서로 다른 변위(disparity)에 의해 입체감을 느끼는 것이다. 그러므로 입체 영화는 두 대의 카메라를 인간의 눈과같이 배치하여 동시에 피사체를 촬영을 하고 이를 인간의 눈에 각각 분리하여 보여주면 똑같은 입체감을 느끼게 된다.

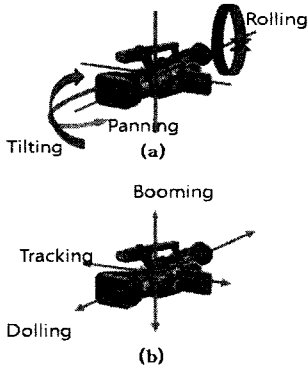


그림 1. 카메라의 움직임  
Fig 1. Camera Motions

동영상은 카메라와 피사체와의 상호작용에 의한 산물이다. 동영상에서의 피사체의 움직임은 피사체 자체가 움직이거나 또는 카메라의 움직임에 의해 발생한다. 카메라의 움직임의 종류를 분류하면 크게 두가지로 나뉜다. 카메라를 일정한 위치에 고정시키고 촬영하는 방법과 카메라를 이동하면서 촬영하는 방법이 있다(그림 1). 카메라를 고정시키고 촬영할 때에는 수평방향으로 회전하는 패닝(Panning)과 수직방향으로 회전하는 틸팅(Tilting), 카메라를 회전하는 롤링(Rolling) 등이 있으며 또한 초점거리를 가변하는 줌잉(Zooming)이 있다. 카메라를 이동 시키면서 촬영하는 방법에는 카메라를 위아래로 이동하면서 촬영하는 붐잉(Booming), 수평방향으로 이동하면서 촬영하는 트래킹(Tracking), 앞으로 이동하는 달링(Dolling)이 있다[8]. 틸팅과 붐잉, 패닝과 트래킹, 줌잉과 달링은 유사한 화면 변화 패턴을 보인다.

움직임이 없는 동영상은 입체 정보가 없으며 이를 무리하게 입체로 구현할 필요는 없다고 판단된다. 패닝/트래킹인 경우 현재 화면과 바로 전 화면을 각각 좌측과 우측 눈에 분리하여 보여 주면 마치 입체용 카메라를 이용하여 촬영한 화면을 보여 주는 것 처럼 입체감을 느낄 수 있다. 즉 단순히 지연된 프레임들 함께 보여 줌에 의하여 입체구현이 가능한 것이다. 틸팅/부밍, 줌잉/달링 및 롤링에 의해 촬영된 화면은 단순 지연된 화면을 보여주는 방법에 의해서는 입체감을 느낄 수 없으며 현기증을 느끼게 한다. 그러므로 기존 프레임들 이용하여 가상의 화면을 합성하는 방법이 제한되고 있다 [7,8].

본 연구에서는 카메라의 움직임에 관계없이 현재프레임과 과거 프레임상의 변위(disparity)를 측정하고 패러렐 카메라 모델에 따른 변위정보를 생성한다. 기준이 되는 왼쪽 영상은 현재 프레임으로 하며 현재 프레임과 바로 이전의 프레임과의

변위정보를 이용하여 합성된 영상은 오른쪽 영상으로 사용한다. 변위는 수렴 반복 알고리즘[11]을 2차원 좌표에서 가능하도록 확장하였다. <그림 2>는 본 연구에서 제안한 방법을 블록다이어그램으로 표현한 것이다.

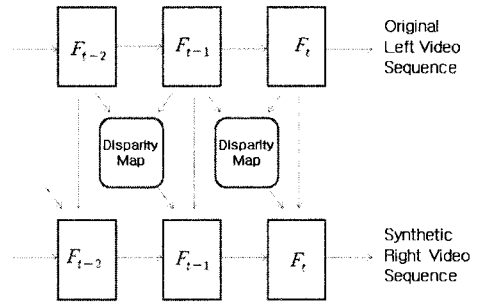


그림 2. 3D변환 블록 다이어그램  
Fig 2. 3D Conversion Block Diagram

장면 전환 등 이전 프레임과 현재 프레임이 연관성이 없는 경우는 MSE 값을 이용하여 판단하며 MSE값이 일정범위를 벗어나는 경우에는 변위맵을 작성하지 않고 왼쪽 영상을 그대로 오른쪽 영상으로 사용한다.

#### IV. 수렴 반복 알고리즘

본 연구에서 제안한 방법은 정확한 변위 맵을 작성할 수 있는가에 따라 입체 영상의 질이 많이 좌우된다. 변위 정보는 두 영상의 정합과정에서 얻을 수 있다. 영상의 정합은 두 영상에서의 점대점 대응관계를 찾는 과정이다[12]. 영상정합은 스테레오 이미지의 거리정보, 의학 영상에서의 변경된 부분 및 물체의 인식에 많이 활용되고 있다.

동영상의 현재 프레임과 이전 프레임은 FPS(Frame Per Sec)가 높을 수록 연산처리는 많지만 이미지의 변화가 적으므로 영상의 정합이 쉬워진다.

영상의 정합법은 여러 가지가 있지만 전통적인 방법은 다음과 같다[13]. 첫 번째 방법은 특징정보추출(Feature-Based)에 의한 매칭방법으로 영상에서 특정의 정보(예 에지 등)를 추출하고 추출된 정보에 의해 매칭 프로세싱을 하므로 연산속도가 빠르고 비교적 신뢰성 있는 변위(disparity) 정보를 구할 수 있다. 그러나 특정정보를 정의하는 것이 매우 어려운 과제이다.

두번째 방법은 두 영상의 계조(Gray Level) 상관도를 이용하는 방법이다. 영상 위의 한 점(m,n)에 대해 상관도를 계산한 후 다른 영상 평면상의 각 점마다 상관도를 구한 후

가장 큰 상관도 값을 갖는 점을 매칭 포인트로 가정한다. 특징정보추출(Feature-Based) 방법은 이산적 변위(Sparse Disparity) 정보를 얻을 수 있으나 이 방법은 연속적 변위(Dense Disparity) 정보를 얻을 수 있는 장점이 있다. 그러나 이 방법은 블록의 크기를 얼마나 하여야 하는가가 문제가 되며, 상관도가 가장 높은 지점이 반드시 대응하는 점이라고 말할 수 있는가? 또한 복수개의 매칭 점들을 어떻게 처리하여야 하는가 등의 문제점이 있다.

본 연구는 정합위치를 찾기 위하여 기존 수렴 반복 알고리즘(9)을 개선하여 정합위치를 찾는다. 이 방법은 상관도에서처럼 두 영상의 계조를 이용하여 2차 에러함수를 정의하고 이 에러함수가 최소가 되도록 정합위치를 반복적으로 찾아간다. 한 영상의 계조를 임의의 2차원 함수  $f(x, y)$ 라고 가정하고 다른 영상에서의 변위된 함수  $f(x - D_x, y - D_y)$ 에서의 불일치 정도  $D_x, D_y$ 를 알아내는 것이다. 즉 대상의 영상이 단순히 기준영상의 변위된 함수로 가정하는 것이다. 그러나 실제로는 단순한 변위가 아니며 카메라의 왜곡과 위치에 따른 다른 함수이다. 2차 에러함수는 (수식 1)과 같이 정의한다.

$$e(d_x, d_y) = \int_{-\infty}^{\infty} (f(x + d_x - D_x, y + d_y - D_y) - f(x, y))^2 dx dy \dots\dots\dots (식 1)$$

이 수식에서 정합된 위치  $d_x = D_x, d_y = D_y$ 인 지점이 에러함수가 최저 값이 되는 지점이다. 위 수식 (1)에서 최소가 되는 지점을 수렴 반복을 통하여 알아내는 것이다. 수렴반복을 통하여 알아내기 위하여 수식(1)을 다음과 같이 수정한다.

$$e(d_x, d_y) = \int_{-\infty}^{\infty} (f(x + d_x + D_u - D_x, y + d_y + D_v - D_y) - f(x, y))^2 dx dy \dots\dots\dots (식 2)$$

여기에서

$$D_u = D_{u-1} - \alpha \nabla e_{k-1}(0,0) \dots\dots\dots (식 3)$$

$$D_v = D_{v-1} - \alpha \nabla e_{k-1}(0,0) \dots\dots\dots (식 4)$$

알고리즘은 수식(2)가 수렴값을 얻을 때까지 반복한다. 즉  $e(d_x, d_y)$  함수가  $e(0,0)$ 에서 최저치가 되도록 반복적으로 수식 (3)(4)를 이용하여  $D_u, D_v$ 를 변경한다. 수식 (3)(4)는 에러함수의 기울기를 이용함으로써 정합위치로 빠

르게 수렴하도록 한다. 만일  $e(0,0) = 0$ 이 된다면 이 지점은 정합된 위치로서  $D_u = D_x, D_v = D_y$ 인 지점인 경우이다. 이 알고리즘은  $e(0,0)$ 에서 미분 값을 얻을 수 있다는 전제 조건이 필요하다.

### V. 변위 맵과 영상합성

입체 영화는 일반적으로 2대의 카메라를 사용하는 스테레오 모델이며 또한 두 대의 카메라가 서로 수평으로 전방을 지향하는 패러렐 카메라 모델이다.

(그림 3)은 핀홀(pin-hole) 패러렐 스테레오 모델을 보인 것이다. 2대의 카메라가 베이스 라인 위에 설치되어 있는 스테레오 카메라 모델은 카메라는 베이스 라인과 수직으로 전방을 향한다. 왼쪽과 오른쪽 카메라를 연결하는 베이스 라인 1/2 지점을 월드좌표의 원점( $O_w$ )으로 지정하고 두 대 카메라 사이의 거리를  $b$  mm라고 가정한다. 카메라의 초점의 길이를  $f$  mm라고 할 때, 카메라의 영상은 카메라의 영상은 베이스 라인으로부터 후방  $f$  mm에 존재하지만 모델링의 편의를 위하여 베이스 라인으로부터 전방  $f$  mm 위치에 있다고 가정한다. 왼쪽과 오른쪽 카메라에 각각 투영된 영상좌표는 카메라 영상 중심(Center  $O_l, O_r$ )을 기준으로 한다.

월드좌표에 있는 한 점  $P(x_w, y_w, z_w)$ 가 왼쪽 카메라  $P_l(u_l, v_l)$ 와 오른쪽 카메라  $P_r(u_r, v_r)$  영상좌표에 각각 투영될때 영상평면의 좌표는 투영변환(Perspective Transform)에 의하여 계산한다.

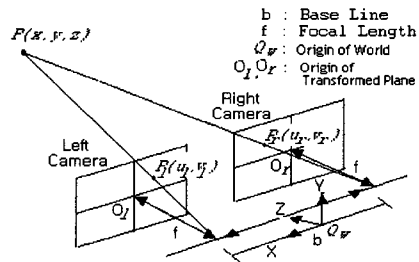


그림 3. 패러렐 카메라 모델  
Fig 3. Parallel Camera Model

수식 (5)은 투영평면의 좌표를 계산한 것이다. 패러렐 카메라 모델은 수식 (5)에서 알 수 있는 바와 같이 수직축의 변위(Disparity)는  $D_v = v_l - v_r = 0$ 으로 월드 좌표의 한



(그림 6)은 동영상의 유형별 입체변환의 예를 보인 것이다. 첫 번째 행의 영상(a)-(d)는 부밍화면이고, 두 번째 행의 영상(e)-(h)는 트래킹 영상이며, 세 번째 행의 영상(i)-(l)영상은 줌인영상이며 마지막 행은 화면상에 부분적으로 로테이션과 줌인이 결합되어 있는 영상이다.

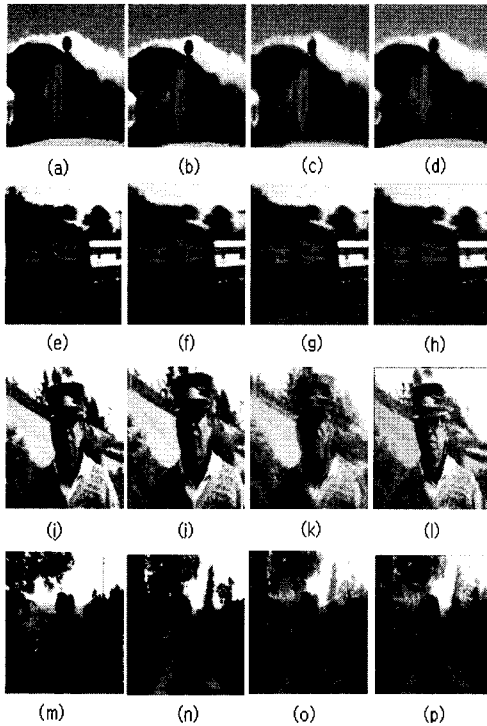


그림 6. 동영상의 유형별 영상합성  
Fig 6. Types of Moving Pictures and Image Synthetics

세로 첫 번째 열(a)(e)(i)(m)은 현재 프레임으로서 기준 영상이며, 두 번째 열(b)(f)(j)(n)은 과거 프레임이고, 세 번째 열(c)(g)(k)(o)는 아무런 처리 없이 단순히 첫 번째열과 두 번째 열의 영상을 중첩한 경우이고, 마지막 열(d)(h)(l)(p)는 본 연구에서 제안한 방법으로 합성한 영상과 현재프레임 영상을 중첩한 경우이다. (c) 영상인 경우 카메라가 위로 움직이는 상황으로 현재 프레임과 과거 프레임은 위아래로 픽셀 변위를 보이지만 합성한 영상(d)는 좌우로만 변위가 되도록 하였다. 트래킹인 (g)영상은 (h) 영상과 별다른 차이가 없으며 단순 지연만으로도 입체화면을 얻을 수 있다. (k)영상은 줌인하는 화면을 중첩한 것으로 현재 프레임과 과거 프레임간 인물 크기가 차이가 나지만 합성한 영상(l)은 좌우 편차만 보인다. (o) 영상인 경우 복잡적으로 줌인과 인물

의 팔이 부분적 회전을 하는 영상을 중첩한 것으로 입체감이 있으나 좀 현기증을 주는데 비하여 합성한 영상 (p)는 좌우 편차만 보이므로 안정된 입체 화면이 재현되었다. 즉 어떤 유형의 동영상이던 간에 스테레오 입체 카메라로 촬영한 것처럼 영상을 합성하여 눈의 피로감이 적고 안정적인 입체영상을 구현하는 것이다.

## VII. 결론

본 연구에서는 일반적인 2차원 비디오 프레임을 3차원 입체 비디오 프레임으로 변환하는 방법이 있어서 연속하는 프레임간 픽셀 변위를 구하고 이 변위 정보를 기반으로 스테레오 카메라로 촬영한 듯한 입체영상을 합성하였다. 픽셀 변위를 구하기 위하여 기존 수렴 반복법을 보완하여 이용하였으며 입체영상 합성을 위하여 수식(7)을 제안하였다. 이 방법에서 가장 중요한 것은 안정적인 픽셀변위를 알아내는 것이며 본 연구에서 사용한 수렴 반복법은 아직 개선의 여지가 많다. 본 연구에서 제안한 알고리즘은 동영상의 어떤 유형이라도 일관성있는 방법에 의하여 합성영상을 만들어 내므로 유형 분류 착오에 의한 문제 점이 없는 페러렐 카메라로 촬영한 듯한 입체 영상을 재현한다. 픽셀 변위를 알아내는 프로세싱 타임이 많이 요구되어 실시간 변환은 어려웠지만 전용 프로세서 등을 개발한다면 실시간 변환도 가능할 것이다.

## 참고문헌

- [1] 신성운, "효율적인 비디오 브라우징을 위한 파노라믹 영상 구축," 한국컴퓨터정보학회 논문지, 제12권, 제1호, 67-72쪽, 2007년, 3월.
- [2] Chen, Chuan Sheng, EP1750458 A1, 3D TV Picture Conversion Device, 02, 2005.
- [3] 이상용, "떨림현상이 없는 입체영상 수신장치," 한국특허 10-0406217, 2003년.
- [4] Ideses, I, "A Method for Generating 3D Video from a Single Video Stream. VMV pp.435-438, 2002.
- [5] T.Okino, "New television with 2D/3D Image Conversion techniques," SPIE vol.2653, Phontonic West, 1995.
- [6] 山河周悟, "2차원 영상을 3차원 영상으로 변환하는 방법", 일본특허, 特開平9-9295, 1997년.

- [7] 김만배, "3차원 입체변환을 위한 MPEG압축데이터에서의 영상처리기법," 한국방송공학회 학술대회, 중앙대학교, 3-8쪽, 1998년. 12월
- [8] Ianir Ideses, "Real Time 2D to 3D Video Conversion," J Real-Time Image Proc, 2, pp.3-9, 2007.
- [9] 한상훈, "특징점의 모션벡터를 이용한 차량 검지시스템 개발," 한국컴퓨터 정보 학회 논문지, 제12권, 제1호, 261-267쪽, 2007년, 3월
- [10] 이희만, "카메라 모션벡터 추출기를 이용한 임베디드 기반 가상현실 시뮬레이터 제어기의 설계," 한국콘텐츠학회논문지, 제6권, 제4호, 98-108쪽, 2006년. 4월.
- [11] 이희만, "스테레오 비전을 이용한 가상환경구현," 한국컴퓨터정보학회논문지, 제9권, 제3호, 95-102쪽, 2004년. 9월.
- [12] A. Ardeshir Goshtasby, "2-D and 3-D Image Registration," Wiley Interscience, 2005.
- [13] Andre Redert, "Correspondence Estimation in Image Pairs", IEEE Signal Processing Magazine, vol. 16, no. 3, pp.29-46, 1999.

## 저자 소개



### 이 희 만

1994년 6월 Texas A&M

Electrical Eng. Ph.D

1996~ 현재 :

서원대학교 컴퓨터공학부

멀티미디어학과 교수