CAVE^{TM}-like 시스템에서 시각 커뮤니케이션 지원을 위한 스테레오 비디오 아바타

(A Stereo Video Avatar for Supporting Visual Communication in a CAVE^{TM}-like System)

이 선 민 † 박 지 영 † 김 명 희 ††
(Seon-Min Rhee) (Ji-Young Park) (Myoung-Hee Kim)

요 약 본 논문에서는 CAVE^{TM}-like 시스템에서 시각 커뮤니케이션 지원을 위한 고화질 스테레오 비디오 아바타 생성 기법을 제안한다. CAVE^{TM}-like 시스템에서는 사용자를 둘러싸고 있는 스크린으로 투사되는 빛의 잦은 변화 때문에 비디오 아바타 생성에 필수적인 사용자 추출이 쉽지 않다. 본 연구에서는 가시광선 차단 필터를 부착한 흑백 카메라로 획득된 적외선 반사 영상을 이용함으로써 스크린 상 빛의 변화를 차단하여 강건하게 사용자를 추출할 수 있도록 하였다. 또한, 사람의 양안차 간격으로 배치한 두 대의 컬러 카메라를 사용하여 삼차원 기하 정보의 재구성 없이 고화질 비디오 아바타를 빠르게 생성하고 입체 디스플레이 하기 위한 양안용 사용자 영상을 획득하였다. 획득된 영상에서 배경을 제거하기 위하여 적외선 반사 영상으로 정의된 실루엣 마스크와의 피팅 알고리즘을 제안한다. 생성된 비디오 아바타 스테레오 영상은 가상공간 내 평면 상에 텍스처 매핑하여 프레임 순차 스테레오 방식를 이용하여 입체 디스플레이할 수 있도록 하였다. 제안된 방식은 기존의 3D 비디오 아바타보다 고화질의 결과를 빠르게 생성할 수 있으며, 2D 기반 방식에서 제공해주지 못하던 입체감을 제공해준다

키워드 : 비디오 아바타, 스테레오 영상 생성, CAVE^{TM}-like 시스템

Abstract This paper suggests a method for generating high quality stereo video avatar to support visual communication in a CAVE^{TM}-like system. In such a system because of frequent change of light projected onto screens around user, it is not easy to extract user silhouette robustly, which is an essential step to generate a video avatar. In this study, we use an infrared reflective image acquired by a grayscale camera with a longpass filter so that the change of visible light on a screen is blocked to extract robust user silhouette. In addition, using two color cameras positioned at a distance of a binocular disparity of human eyes, we acquire two stereo images of the user for fast generation and stereoscopic display of a high quality video avatar without 3D reconstruction. We also suggest a fitting algorithm of a silhouette mask on an infrared reflective image into an acquired color image to remove background. Generated stereo images of a video avatar are texture mapped into a plane in virtual world and can be displayed in stereoscopic using frame sequential stereo method. Suggested method have advantages that it generates high quality video avatar faster than 3D approach and it gives stereoscopic feeling to a user 2D based approach can not provide.

Key words : Video Avatar, Stereo Image Generation, CAVE^{TM}-like system

1. 서 론

최근 개인용 컴퓨터 성능이 급속하게 향상되어 저비용으로 대용량 정보를 빠르게 처리할 수 있게 됨에 따라 과학적 가시화, 건축 및 자동차 디자인 등과 같이 입체 및 실물 크기 가시화를 통하여 효과적인 결론 도출이 가능한 분야에 대한 관심이 높아지고 있다. 이와 같은 요구를 충족시키기 위하여 전 세계적으로 많은 대학이나 연구소에서 CAVE^{TM}-like 시스템과 같은 대형의 몰입형 가상환경을 구축하고 이를 이용한 다양한 연구를 수행하고 있다. 뿐만 아니라, 초고속 통신망이 널리 보급됨에 따라 개별 가상환경을 네트워크로 연동한 공

† 학생회원 : 이화여자대학교 컴퓨터학과
blue@ewhain.net
lemie@ewhain.net
†† 종신회원 : 이화여자대학교 컴퓨터학과 교수
mhkim@ewha.ac.kr
논문접수 : 2005년 1월 24일
심사완료 : 2006년 3월 15일

유가상환경을 구축하려는 시도 또한 늘고 있다. 공유가상환경에서는 개별 가상환경 내에 존재하는 사용자간의 원활한 의사 소통이 중요하며, 특히 사용자의 위치, 방향 및 행위 정보 등을 상대방에게 효과적으로 보여주기 위한 시각 커뮤니케이션(visual communication) 지원 기법이 필수적으로 요구된다.

비디오 아바타(video avatar)는 카메라로 입력 받은 사용자 영상을 가상세계에 투영하여 가상객체와 합성하여 보여주는 기법이며, 시각 커뮤니케이션을 효율적으로 지원하기 위한 수단으로 효과적으로 활용될 수 있다[1]. CAVE™-like 시스템과 같은 가상현실 환경에서 비디오 아바타를 이용하여 보다 원활한 시각 커뮤니케이션을 제공하기 위해서는 다음 두 가지 사항을 만족시켜야 한다. 우선, 조명 조건이 일정하지 않은 환경에서 획득된 영상으로부터 배경을 제거하고 사용자를 강건하게 추출할 수 있어야 한다. 일반적인 배경제거(background subtraction) 기법은 학습 전후의 배경이 동일함을 전제로 하지만 CAVE™-like 시스템의 경우, 사용자를 둘러싸고 있는 스크린상에 디스플레이 되는 콘텐츠가 변함에 따라 배경도 바뀌게 되므로 사용자 분할이 쉽지 않다. 기존의 비디오 아바타 관련 연구에서는 강건한 사용자 추출을 위하여 블루 스크린을 이용하는 등 영상 획득에 최적화 된 별도의 공간을 마련하는 경우가 대부분이었다[1-7]. 그러나 영상 획득을 위한 별도의 환경상에 있는 사용자의 모습만 단방향으로 전송되기 때문에 양방향 커뮤니케이션이 불가능하다. 따라서 별도의 공간이 아닌 디스플레이 환경 내에 존재하는 사용자 영상을 직접 획득하고 추출할 수 있어야 한다. 다음으로, 가상세계를 경험하고 있는 사용자의 몰입감을 저해하지 않고 현실감을 증진시키기 위하여 고화질의 비디오 아바타를 실시간으로 입체 디스플레이 할 수 있어야 한다. 비디오 아바타는 삼차원 기하 정보 생성 여부에 따라 크게 2D 및 3D로 나누어 볼 수 있다[1]. 2D 비디오 아바타는 입력 영상을 그대로 사용하기 때문에 빠르고 결과 영상의 질이 좋은 반면, 3D 비디오 아바타는 기하정보 재구성에 걸리는 시간이 길고, 텍스처 매핑 시 발생되는 입력 영상의 변형에 의해 비사실적으로 보이기 쉽다. 고화질의 비디오 아바타를 실시간으로 생성하기 위해서는 원영상에 변형을 가하지 않는 2D 방식이 바람직하지만, 이 경우 깊이에 따른 원근감 및 객체 자체의 부피감을 표현할 수 없다는 단점이 있다.

본 논문에서는 CAVE™-like 시스템과 같이 정적인 배경을 보장할 수 없는 대형 프로젝션 기반 가상환경 내에 있는 사용자의 고화질 비디오 아바타를 입체 디스플레이하여 시각 커뮤니케이션을 증진시키기 위해 필요한 기술 요소들을 제안한다. 스크린으로 투사되는 빛의 변화를 차단하고 사용자 영역을 강건하게 추출하기 위하여 적외선 반사 영상을 이용한 관심 영역 추출을 제안하며, 정의된 사용자 실루엣을 이용하여 텍스처 영상에서의 배경을 제거하기 위한 실루엣 및 텍스처 영상간 피팅 알고리즘을 제안한다. 본 연구에서는 기존의 2.5D나 3D로 모델의 재구성 하는 방식과는 달리 원영상에 변형을 가하지 않는 2D 기반 비디오 아바타 생성 방식을 채택함으로써 고화질의 결과 영상을 생성할 수 있도록 하였으며, 별도의 영상획득 공간이 아닌 CAVE™-like 시스템 내에 있는 사용자 영상을 직접 획득하도록 한다는 점에서 기존의 연구와의 차별성을 갖는다.

본 논문의 구성은 다음과 같다. 2장에서는 비디오 아바타에 관련된 국내외 연구 동향 및 가상환경에서 주로 사용되는 입체 렌더링 방식에 대하여 설명한다. 3장에서는 본 논문에서 실험한 하드웨어 환경에 대하여 기술하며, 4장에서는 비디오 아바타 스테레오 영상 생성 방법에 관하여 설명한다. 5장에서 실험 결과에 대하여 기술하고 마지막으로 6장에서는 결론 및 향후 연구 방향에 대하여 제시한다.

2. 관련 및 배경 연구

2.1 비디오 아바타

비디오 아바타는 크게 2D-, 2.5D-, 3D 방식으로 나누어 볼 수 있다[1]. 2D 비디오 아바타는 가상공간 내의 특정 위치에 존재하는 평면상에 사용자 영상을 텍스처 매핑하여 보여주는 형태이다. 이는 매우 간단하고 고화질의 결과 영상을 생성할 수 있지만 깊이를 구할 수 없기 때문에 입체감을 제공해줄 수 없다는 단점이 있다. 반면, 3D 아바타는 다수의 카메라를 이용하여 여러 각도에서 사용자의 모습을 촬영하고 이를 삼차원 모델로 재구성하여 보여주기 때문에 임의의 시점에서 입체감 있는 사용자 모습을 제공할 수 있지만 모델 재구성에 걸리는 시간이 방대하다는 단점이 있기 때문에 실시간 상호작용형 어플리케이션에 사용되기 어렵다. 2.5D 비디오 아바타는 두 가지의 단점을 보완하고자 제안된 방법으로 스테레오 카메라를 이용하여 깊이값을 추정하고, 특정 시점에서의 사용자의 모습을 삼차원 모델로 재구성하여 보여주는 형태이지만 시점이 바뀔 경우에는 정상적인 삼차원 모델을 제공할 수 없다.

실시간 입력 영상을 이용한 비디오 아바타 생성에 관한 국내외 연구는 다음과 같다. 한국과학기술 연구원(KIST) 등에서는 가상환경에서 다수의 실시간 비디오 아바타를 생성하기 위한 연구를 수행하였다[2]. 이 연구에서는 사용자 영상을 촬영하는 카메라와는 별도로 보조 카메라를 설치하여 사용자의 움직임을 추적함으로써

비디오 아바타가 보다 자연스러운 방법으로 가상공간 상의 다른 객체들과 합성될 수 있는 방법을 제안하였다. 이 연구에서 제안된 비디오 아바타는 깊이값을 갖지 않기 때문에 2D 비디오 아바타로 분류될 수 있으며, 사용자의 움직임 추적이 가능하다는 강점을 갖는다. 광주과학기술원(KJIST)에서는 네트워크 가상환경을 위한 단순화 된 2.5D 비디오 아바타 생성 기법을 제안하였다[3,4]. 이 연구에서는 아바타 모델을 구성하는 데이타 양의 실시간 단순화 알고리즘을 제안하여 제한된 네트워크 대역폭 내에서 사실감 있는 비디오 아바타를 생성할 수 있도록 하였지만, 디스플레이 환경 내에 있는 사용자를 대상으로 하지 않았다. T. Ogi et al.은 몰입형 가상환경에서의 커뮤니케이션 수단으로써 비디오 아바타를 제안하고 있다[5]. 이 연구에서는 비디오 아바타를 플레인 모델(plane model), 깊이 모델(depth model), 복셀 모델(voxel model), 얼굴 모델(face model)로 분류하고 각 기법의 장단점을 서술하여 응용 분야의 목적에 적합한 방법을 선택할 수 있는 근거를 제시하였다. V. Rajan et. al은 네트워크 가상환경에서 사용자의 머리 모델을 3차원으로 재구성하기 위한 방법을 제안하였다[6]. 이 논문에서는 별도의 트래커 정보를 이용하고 있으며, 시점 종속 텍스처(view dependent texturing) 기법을 통하여 현실감 및 몰입감을 높일 수 있도록 하였다. K. Cheung et al[7]에 의해 제안된 방법은 기존의 SFS (shape from silhouette) 기법을 개선하여 보다 빠른 사용자의 복셀 모델을 생성하여 실시간 어플리케이션에 최적화 시키는 연구를 수행하였다. 그러나 CAVE^{TM}-like 시스템과 같이 어둡고 한정적인 공간상에 설치되어 있는 가상현실 환경에서는 양질의 텍스처를 생성하기 위해 필요한 충분한 개수 및 다양한 각도의 카메라를 설치하기가 어렵고, 칼라 카메라만을 이용해서는 사용자 실루엣을 추출하기 어렵다. 또한 영상의 생성뿐만 아니라 생성된 영상의 전송속도를 고려해야 할 때 이 방법을 시각 커뮤니케이션 지원을 위해 활용하기에는 현실적으로 무리가 있다.

관련 연구를 살펴볼 때 실시간 카메라 영상을 이용하여 사용자의 모습을 재구성하고자 하는 시도는 다각도로 이루어지고 있지만 대부분의 경우 사용자 분할을 원활하게 하기 위하여 별도의 공간상에서 영상을 획득하고 있으며 가상환경 내에 있는 사용자 영상을 직접 이용한 경우는 드문 편이다. 또한, 임의의 시점에서 입체감 있는 사용자 영상을 제공하기 위하여 2D 보다는 3D 방식을 택하는 경우가 많지만 고화질의 비디오 아바타를 생성하기 위해서는 입력 카메라의 개수를 늘려야하며 이에 따른 계산량 및 모델 자체의 용량 증가로 인하여 실시간으로 상대측에 사용자 영상을 전송해주기 어려운 실정이다.

2.2 가상 환경을 위한 사용자 영상 획득

어둡고 조명 조건이 일정하지 않은 프로젝션 기반 가상환경에서 사용자 영상을 효과적으로 촬영하기 위해 제안된 연구는 다음과 같다. M.Gross et al.[8]은 실사 기반 사용자 영상을 이용한 텔레프레센스 제공에 관한 연구를 수행하고 있지만, 스크린에 투사되는 빛을 차단하기 위하여 기존의 하드웨어를 모두 변형해야 하므로 일반적인 몰입형 가상환경에 적용하기엔 무리가 있다. 서로 다른 카메라를 이용한 사용자 영역 정의 및 색상 정보 추출에 관한 연구는 P. Debevec et al.[9], Yasuda, K et al.[10], S.Y. Lee et al.[11]등에 의해 진행되어 왔지만, 두 카메라 사이의 빔 스플리터(beam splitter) 혹은 반투명 거울(semi-transparent mirror)을 설치하여 빛의 일부는 투과 시키고, 일부는 반사시킴으로써 두 카메라에 같은 영상을 동시에 입력시키는 방식을 이용한다. 이 방법은 카메라의 위치 조정이 까다롭고, 빛의 일부가 손실되어 카메라로 입력되기 때문에 일반적으로 프로젝션 기반 가상환경이 어둡다는 점을 감안할 때 효과적이지 못하다.

이와 같이 비디오 영상으로부터 실사 기반 사용자 영상을 추출하기 위한 연구가 다양하게 진행되고 있지만, 별도의 하드웨어 변형 없이 조명 조건이 일정하지 않은 어두운 환경에서의 사용자 추출에 관한 연구는 드문 실정이다.

2.3 스테레오스코픽(stereoscopic) 디스플레이

프레임 순차 스테레오는 컴퓨터에 의해 생성된 삼차원 객체들을 이차원 모니터 혹은 스크린 상에 입체로 보여주기 위하여 가장 널리 이용되는 방법이다. 사람이 사물을 입체적으로 인지하는 것은 좌·우 양 눈이 가로 방향으로 평균 65mm 떨어져 있기 때문에 서로 다른 2차원 영상을 보게 되고, 이 영상이 망막을 통해 뇌로 전달되면 뇌에서 이를 융합하여 본래의 3차원 영상의 깊이감을 재생하기 때문이다. 프레임 순차 스테레오 방식은 이러한 원리를 이용하여, 삼차원으로 정의되는 가상세계를 바라보는 시점을 조정하여 양안에 해당하는 영상을 각각 생성하여 빠르게 교차 디스플레이하고, 사용자가 셔터글래스 등을 통해 각 눈에 해당하는 영상만 볼 수 있도록 함으로써 입체감을 제공한다.

양안 영상은 그림 1에서 보는 것과 같이 양안차(binocular disparity) 간격으로 나란히 배치된 두 대의 카메라 영상 평면상에 투영되어 생성될 수 있으며, 삼차원 공간상의 임의의 점은 두 카메라의 물리적인 위치 차이에 의해 각 영상의 서로 다른 좌표에 투영된다. 이때, 삼차원 공간상의 위치가 카메라에서 가까우면 두 영상간의 양안차는 작아지고, 멀어질수록 커지게 되므로

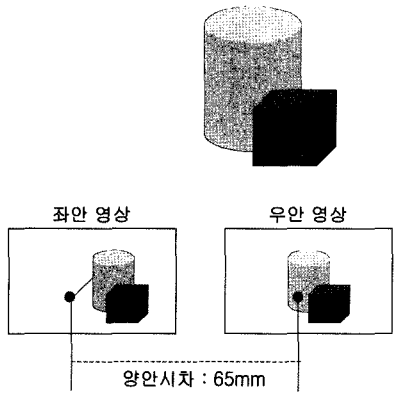


그림 1 두 카메라에 의한 양안 영상 생성 예

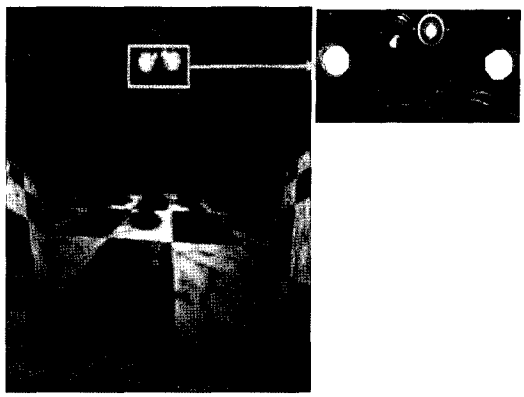


그림 2 하드웨어 셋업 : CAVE™-like 시스템 내 설치된 두 쌍의 실루엣-텍스처 카메라, 적외선 광원 및 주변광원

객체간 위치뿐 아니라 하나의 객체 내 볼륨도 표현 가능하다. 따라서 카메라 두 대를 이용하여 생성된 양안 영상을 프레임 순차 스테레오 방식의 환경에 맞춰 교차 디스플레이 하게 될 경우 삼차원 정보 재구성 없이도 객체의 입체감을 표현할 수 있다.

기존의 2D 비디오 아바타는 배경 제거 등을 통하여 추출된 한 장의 사용자 영상을 가상환경 내에 존재하는 평면상에 텍스처 매핑하여 보여준다. 따라서 라이브러리 및 그래픽 카드에서 제공하는 스테레오 기능을 사용하게 되더라도, 인체의 볼륨에 관계 없이 가상세계 내 존재하는 평면의 깊이에 따라 동일한 양안차가 생성되기 때문에 입체감을 줄 수 없다. 반면, 제안하는 스테레오 영상을 이용한 비디오 아바타는 CAVE™-like 시스템 내에 양안차 간격으로 나란히 배치된 두 대의 카메라로 획득된 사용자 영상 각각을 양안용 텍스처로 제공한다. 따라서 사람의 신체 부위별 깊이에 따라 서로 다른 양안차를 갖는 두 영상이 생성되며, 이를 가상세계 내에 교차 디스플레이하고 동기를 맞춰줄 경우 사용자가 비디오 아바타의 부피감을 인지할 수 있게 되므로 삼차원 기하정보를 재구성하지 않고도 입체감을 제공할 수 있다.

3. 하드웨어 구성

본 연구에서는 가상환경 내에 입체 디스플레이 될 비디오 아바타 스테레오 영상을 생성하기 위하여 흑백 카메라 두 대와 칼라 카메라 두 대를 동시에 이용하였다. 흑백 카메라(이하, 실루엣 카메라)는 사용자의 영역을 정의하기 위하여 사용되며 칼라 카메라(이하, 텍스처 카메라)에서 배경을 제거하기 위해 이용된다. 양안용 스테레오 영상을 생성하기 위하여 그림 2에서 보는 것과 같이 텍스처 카메라 두 대를 사람의 양안차 간격에 해당하도록 배치하였고, 양 옆에 실루엣 카메라를 각각 배치하였다. 사용된 카메라는 Point Grey Research사의 Dragonfly™(해상도 : 640x480, 320x240)이며, firewire 카드 2개를 이용하여 하나의 컴퓨터에 동시 연결함으로써 보다 쉽게 각 영상을 동기화 시키도록 하였다. 사용된 컴퓨터는 Dell Dimension 8300(Intel(R) Pentium(R) 4 CPU 3.00 GHz x 2, 2GB RAM, NVIDIA GeForce FX 5200)이다.

프로젝션에 의해 일정한 조명 조건을 보장할 수 없는 가상환경에서 보다 강건하게 사용자 영역을 정의하기 위하여 적외선 소스 및 가시광선 차단 필터 사용하였다. 이를 이용하면, 스크린 상의 가시 광선 변화는 차단하고, 적외선 광원에 반사되는 사용자 영상만이 카메라에 입력되기 때문에 보다 효과적으로 사용자 영역을 정의할 수 있다. 실루엣 카메라에 700nm이하 파장인 가시광선을 차단하는 long pass filter를 부착하였으며, 715nm 대의 파장을 방출하는 적외선 광원을 설치하였다. 특히, 사용자의 머리 부분나 다리 부분까지도 적외선이 골고루 분산되도록 하기 위하여 전면 스크린 상단 부분과 하단에 두 개씩 배치하였다. 이 밖에도 사용자 영상 획득에 필요한 빛을 제공하기 위한 주변 광원을 설치하였으며, 이 때 적외선은 방출하지 않고 가시광선만 방출하는 광원을 선택하여 사용자 추출에 방해가 되지 않도록 하였다.

4. 스테레오 비디오 아바타

물리적으로 수평선 상에 나란히 배치된 두 대의 카메라에서는 객체의 깊이나 볼륨에 따라 x축 상으로 서로 다른 디스패러티를 갖는 영상이 생성된다. 따라서 사람의 양안차 간격으로 배치된 카메라에서 생성된 영상을 평면상에 차례로 교차 투사하고 사용자가 셔터글래스를 통하여 각 눈에 해당되는 영상만을 볼 수 있게 동기화

하면 객체의 삼차원 기하정보를 재구성하지 않고도 입체가시화 할 수 있다. 본 연구에서는 이러한 원리를 적용하여 그림 2에서 제시했던 것과 같이 양안차 간격으로 배치된 두 대의 칼라카메라로 사용자 영상을 획득하고 배경을 제거함으로써 입체 디스플레이 가능한 비디오 아바타 스테레오 영상을 생성하였다. 이를 위하여 우선, 두 쌍의 실루엣 및 텍스처 카메라로부터 동기화 된 4장의 영상을 획득하고, 적외선 반사영상을 이용하여 실루엣 마스크를 생성한다. 생성된 마스크는 4.3절에서 제안하는 실루엣 피팅 알고리즘을 이용하여 텍스처 영상에 피팅시킴으로써 배경을 제거할 수 있으며 이를 통하여 두 장의 비디오 아바타 스테레오 영상을 생성하고 이를 가상세계에 합성하여 입체 디스플레이 한다. 이를 위한 파이프라인 및 예상 결과물은 그림 3과 같다.

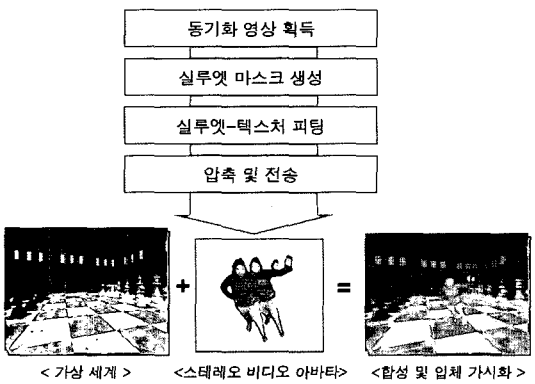


그림 3 두 쌍의 실루엣-텍스처 영상을 이용한 비디오 아바타 스테레오 영상 생성 수행과정 및 가상세계와의 합성 디스플레이 예

4.1 동기화 영상 획득

동일 시점에서의 사용자 영상을 추출하여 2D 비디오 아바타를 생성하기 위해서는 네 대 카메라의 동기화가 필수적이다. 동일한 firewire 카드에 연결되어 있는 두 카메라는 자동적으로 동기화 된다. 따라서 상대적으로 동기화가 더 중요한 한 쌍의 실루엣-텍스처 카메라를 동일한 firewire 카드에 연결하였으며, 서로 다른 firewire 카드에 연결되어 있는 카메라간 동기는 Point Grey Research 사의 flycapture 라이브러리 제공 함수를 이용하였다.

4.2 실루엣 마스크 생성

사용자를 추출하기 위해서는 배경 제거(background subtraction)가 필수적이며, 일반적으로 배경 제거에 사용되는 기법은 크로마 키잉(chroma keying)과 같이 배경색이 동일하거나 조명 조건이 일정함을 가정한다. 그러나 CAVE™-like 시스템에서는 스크린으로 프로젝션되는 빛의 변화로 인하여 위의 조건을 만족시키기 어렵다. 따라서 본 연구에서는 적외선 반사영상(infrared reflective image)을 이용하여 이러한 빛의 변화를 차단하고 사용자 윤곽을 효과적으로 정의할 수 있도록 하였다[12].

4.2.1 적외선 반사 영상

적외선 반사 영상은 흑백 카메라에 가시광선 차단 필터를 장착하고 적외선 광원으로부터 물체에 반사된 빛만 흡수하여 생성되는 영상이다. 스크린에 투사되는 빛은 대부분 0-700nm 대의 가시광선이다. 따라서 적외선 반사 영상을 이용하면, 프로젝션에 의해 변화되는 빛은 차단되고 사용자 몸에 반사된 적외선만 감지해 낼 수 있기 때문에 사용자를 둘러싼 배경이 정적이라고 가정할 수 있으며 기존의 배경제거 기법을 그대로 적용시킬 수 있다. 따라서 조명 조건이 일정하지 않은 프로젝션 기반 가상환경에서의 배경 제거에 매우 효과적으로 활용될 수 있다. 그림 4는 광원으로부터 방출된 적외선이 물체에 부딪혀 카메라로 입력된 결과 영상 예를 보여준다. 일반 흑백 카메라 영상과 비교해 보았을 때, 가시광선 차단 필터 및 적외선 광선을 동시에 설치하고 획득한 영상에서는 프로젝션에 의해 변화하는 가시광선이 모두 차단되어 나타나지 않고 있음을 알 수 있다.

적외선 반사 영상에서는 사용자 이외의 배경이 정적이라고 가정할 수 있으며 다음의 알고리즘[13]을 적용하여 배경을 제거하고 사용자 실루엣 마스크를 정의하였다. 사용된 알고리즘은 배경습득, 전경추출 및 후처리 단계로 이루어진다.

배경습득 단계에서는 사용자가 존재하지 않는 배경영상의 n개 연속 프레임을 저장하고, 각 픽셀위치의 n개 값에 대한 평균, 표준편차를 계산한다.

배경분리 단계에서는 사용자를 포함한 연속 영상이 입력되며 각 픽셀값과 배경습득 단계에서 계산된 동일

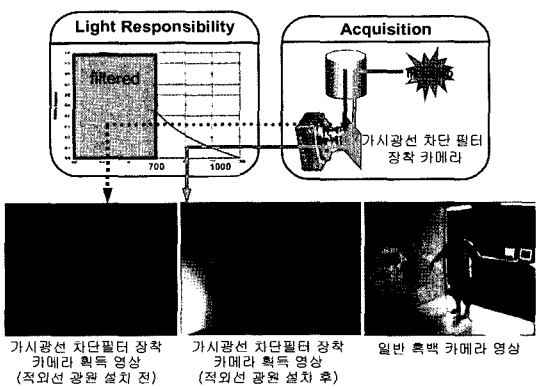


그림 4 적외선 광원 및 가시광선 필터를 이용한 적외선 반사 영상 생성 예

4.2.2 배경 제거

위치상의 통계값을 식 (1)을 이용하여 비교함으로써 변화 여부를 판별한다. 식 (1)에서 p는 해당 픽셀, v는 픽셀값, μ와 σ는 각각 습득영상에 대해 사전에 계산된 각 픽셀위치의 평균 및 표준편차이고 k는 상수이다. 입력영상의 픽 셀값(v_p)과 습득 영상들의 평균값(μ_p)의 차이가 습득 영상들의 표준편차(σ_p)의 k배 보다 크면 해당 픽셀은 변화가 일어났다고 간주되어 사용자 영역에 포함된다.

$$\text{If } (v_p - \mu_p) > k * \sigma_p \text{ then } p \text{ is foreground} \quad (1)$$

후처리 단계에서는 팽창(dilation)과 미디언 필터(median filter)를 적용하여 배경분리 단계를 거친 영상의 잡음을 제거하고 사용자 실루엣을 보다 자연스럽게 나타나도록 한다.

그림 5는 적외선 반사 영상에 위의 알고리즘을 적용하여 추출된 사용자 실루엣 마스크 및 이에 대한 이진 영상이다.

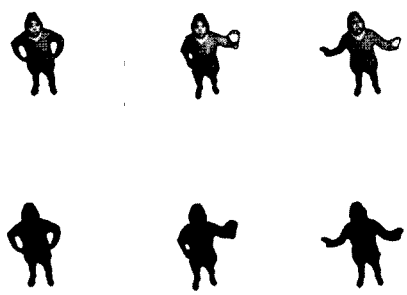


그림 5 사용자 실루엣 마스크 : (a) 적외선 반사 영상을 이용한 사용자 추출 결과, (b) (a)의 이진(binary) 영상

4.3 실루엣 피팅

적외선 반사 영상 내에 정의된 사용자 실루엣 마스크를 텍스처 영상에 적합하여 배경이 제거된 사용자 영상을 생성하기 위해서는 캘리브레이션으로 획득된 카메라 내·외의 파라미터와 디스패러티를 이용하여 두 영상간의 변환 행렬을 구해야 한다. 이 때, 그림 6에서 보는 것과 같이 실루엣 및 텍스처 영상을 각각 렉티피케이션 시켰다고 가정하여 두 영상 내 대응점 검색을 일차원(x축)으로 감소시킨다. x축 상으로의 디스패러티는 두 실루엣 마스크 상에 정의된 대응점을 이용하여 추정된 깊이값, 두 카메라간의 베이스라인, 초점거리를 이용하여 계산할 수 있으며 이를 기반으로 구해진 두 영상간의 변환 행렬을 실루엣 마스크에 적용하여 텍스처 영상에서의 사용자 영역을 정의하고 배경을 제거할 수 있다.

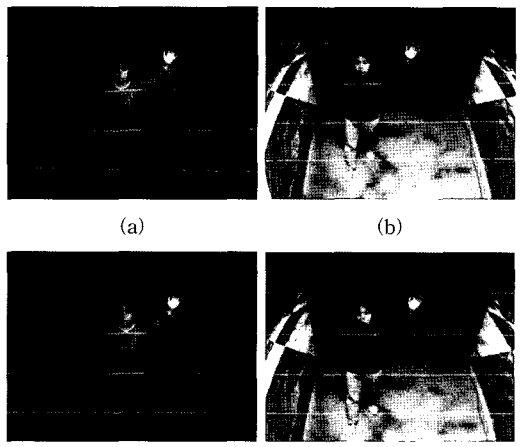


그림 6 실루엣 및 텍스처 영상 렉티피케이션을 통한 대응점간 에피폴라 라인 : (a)(b) 렉티피케이션 전 실루엣 및 텍스처 영상 (c)(d) 렉티피케이션 후 실루엣 및 텍스처 영상

이 때, 실루엣 및 텍스처 영상 피팅에 필요한 사용자의 삼차원 공간상의 깊이값은 하나라고 가정하였다.

4.3.1 카메라 캘리브레이션

양안 영상을 얻기 위해서는 삼차원 공간상의 깊이값 추정이 필수적이며, 이를 위해서는 카메라의 물리적인 위치, 방향 및 내부적인 정보가 필요하다. 본 연구에서는 카메라 내·외부 파라미터를 구하기 위하여 Tsai[14]가 제안한 방법을 이용하였다. 외부 파라미터에는 카메라 및 세계 좌표계간의 이동 및 회전정보가 포함되며, 내부 파라미터는 카메라 좌표계상의 점을 실제 이미지상에 투영하기 위해 필요한 정보인 카메라 초점거리, 이미지 평면 중심, 비틀어짐 값이 포함된다. 카메라 내·외부 파라미터를 구하면 삼차원 공간상의 좌표가 투영되는 2차원 이미지상의 좌표값을 얻기 위한 투영행렬(projection matrix)을 아래와 같이 구할 수 있다. 이 때 P는 투영행렬, A는 카메라 내부파라미터 행렬, R은 회전행렬, T는 이동행렬, C는 카메라 중심을 나타낸다.

$$P = A[R|T] = AR[I|-C]$$
$$= \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix} = \left[\begin{array}{c|c} q_1^T & q_{14} \\ q_2^T & q_{24} \\ q_3^T & q_{34} \end{array}\right] = [Q|\bar{q}] \quad (1)$$

식 (1)에서 정의된 투영 행렬을 이용하여 식 (2)와 같이 카메라 중심 C(COP)를 구할 수 있으며 이는 양안차 계산에 필요한 베이스라인의 길이를 구하는데 이용된다.

$$C = -Q^{-1} * \bar{q} \quad (2)$$

4.3.2 실루엣-텍스처 영상 렉티피케이션

한 쌍의 스테레오 영상의 에피폴라 라인이 특정축(주로 x축)을 기준으로 했을 때 동일선상에 서로 평행하게 놓일 수 있도록 영상 평면을 조정하는 과정을 렉티피케이션[15]이라 한다. 렉티피케이션을 수행하면 두 영상 내 대응점 검색 범위가 이차원에서 일차원으로 줄어들기 때문에 검색 시간을 단축시킬 수 있다는 장점이 있다. 렉티피케이션 영상은 기준 축을 중심으로 두 카메라 중심이 나란히 놓이도록 카메라를 회전하여 구해지는 새로운 투영행렬에 의해 재정의된다. 이 때 두 카메라의 중심은 바뀌지 않으며, 두 카메라가 동일한 초점 거리를 갖도록 내부 파라미터를 임의로 조정할 수 있다. 기존의 투영 행렬을 식 (1)에서와 같이 $\bar{P}_o=[Q_o|\bar{q}_o]$, 렉티피케이션에 의해 구해지는 새로운 투영행렬을 $\bar{P}_n=[Q_n|\bar{q}_n]$라고 정의했을 때, $T_1=Q_nQ_o^{-1}$로 정의되는 호모그래피(homography)를 원영상에 적용함으로써 렉티피케이션 영상을 얻을 수 있다.

본 연구에서 양안영상의 새로운 투영 행렬을 $\bar{P}_{nl}=A[R|-RC_l]$, $\bar{P}_{nr}=A[R|-RC_r]$라고 정의했을 때 A는 실루엣 카메라의 내부 파라미터, C_l, C_r은 두 카메라의 중심이며 회전행렬 R은 식 (4)와 같이 정의된다. 새로운 x축에 해당하는 r1은 두 카메라 베이스라인과 평행하게 정의하며, y축을 정의하는 r2는 실루엣 카메라의 초기 z축 및 새로운 x축과 orthogonal 하게 정의한다. 새로운 z축을 정의하는 r3는 r1과 r2의 orthogonal 한 벡터로 정의한다.

$$R=\begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix}$$

$r_1=(C_l-C_r)/\|C_l-C_r\|$, $r_2=k\Lambda r_1$, $r_3=r_1\Lambda r_2$, k는 임의의 상수 (3)

4.3.3 실루엣-텍스처 영상간 디스패러티

실루엣 및 텍스처 영상을 렉티피케이션 시켰다고 가정할 경우, 두 영상 간의 모든 대응점은 x 축상으로의 디스패러티를 갖게 되며 이를 계산하여 실루엣 마스크를 이동 변환하면 텍스처 영상에서 사용자 영역과 배경을 쉽게 구분해 낼 수 있다. 디스패러티는 두 카메라 중심(COP)간의 거리, 카메라 초점거리, 깊이값을 이용하여 식 (4)와 같이 구할 수 있다.

$$disparity=\frac{b*f}{d} \quad (4)$$

(b :베이스라인 길이, f: 초점거리, d : 깊이값)

디스패러티 추정에 필요한 필요한 삼차원 공간상에서 사용자의 깊이를 계산하기 위하여 식 (5)에서 보는 것과 같이 두 실루엣 마스크의 중심점 C(centroid)를 대응점으로 가정하였다. 이 때, n은 실루엣 마스크로 정의된 픽셀 수, xi, yi는 실루엣 마스크 내 i번째 픽셀의 좌표이다. 구해진 두 개의 대응점을 Triangulation[16] 알고리즘에 적용하여 세계좌표계(world coordinate)상 사용자의 깊이를 구하며, 칼만 필터[17]를 적용하여 잡음에 의한 대응점 검색의 오류를 최소화 하도록 하였다.

$$C=\left(\frac{\sum_{i=1}^{n}x_i}{n},\frac{\sum_{i=1}^{n}y_i}{n}\right) \quad (5)$$

4.3.4 실루엣 및 텍스처 영상간 변환 행렬

4.3.2절 및 4.3.3절에서 기술한 바와 같이 실루엣 및 텍스처 영상을 각각 렉티피케이션 시킨 후 계산된 디스패러티 만큼 실루엣 영상을 이동 변환하면 텍스처 영상의 배경을 마스킹 할 수 있지만(그림 7(a)->(b)->(c)->(d)) 렉티피케이션 과정에서 영상 왜곡이 발생된다. 따라서 영상을 직접 렉티피케이션 하지 않고 렉티피케이션에 필요한 호모그래프(homograph) 및 디스패러티 만큼 이동 변환의 역변환 행렬을 이용하면 그림 7(a) ->(e)->(d)와 같이 텍스처 원영상을 직접 마스킹 할 수 있기 때문에 고화질의 비디오 아바타를 생성할 수 있다.

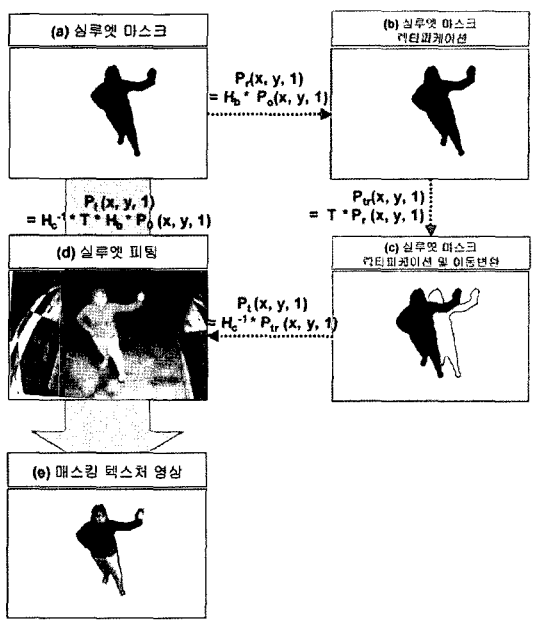


그림 7 실루엣 마스크를 이용한 텍스처 배경 마스킹. Po (x,y,1)은 실루엣 마스크내 임의 픽셀 위치, Ho는 실루엣 영상 렉티피케이션을 위한 호모그래피, Pr(x,y,1)은 실루엣 영상의 렉티피케이션 후 임의 픽셀 위치, Ptr(x, y, 1)은 렉티피케이션 된 실루엣 마스크의 이동 변환 후 픽셀 위치, T는 디스패러티 만큼 이동 변환 행렬, Hc는 텍스처 영상의 렉티피케이션을 이용한 호모그래피, Pt(x,y,1)은 렉티피케이션 및 이동변환 후의 실루엣 마스크의 픽셀 위치

5. 구현 및 결과

제안한 방법에 의해 생성된 스테레오 비디오 아바타 영상은 압축되어 렌더링 서버에 전송된다. 렌더링 서버에서는 전송된 좌안 영상과 우안 영상이 프레임 순차 스테레오 방식에 동기화되어 디스플레이 되도록 해주어야 한다. 본 연구에서는 OpenGL Performer 기반의 blue-c API[18]를 이용하여 압축 전송된 양안 영상을 디코딩하고 가상세계 내에 존재하는 평면상에 텍스처 매핑하고, 각 눈에 맞는 영상을 제공하기 위한 동기화를 맞춰주었다.

표 1에서는 영상의 해상도가 640×480, 320×240일 경우에 각 파이프라인 별로 수행 시간을 측정한 결과이다. 320×240일 경우, 비디오 아바타 스테레오 영상을 만드는 데 소요되는 시간은 약0.04초이므로 약 25frame/sec로 비디오 아바타를 가상환경 상에 디스플레이 할 수 있으며, 640×480 해상도인 경우, 약 6frame/sec이다. 따라서 높은 해상도의 영상을 이용하고자 할 경우에는 프로세서를 추가하여 양안을 위한 영상을 따로 생성하는 것이 바람직하다.

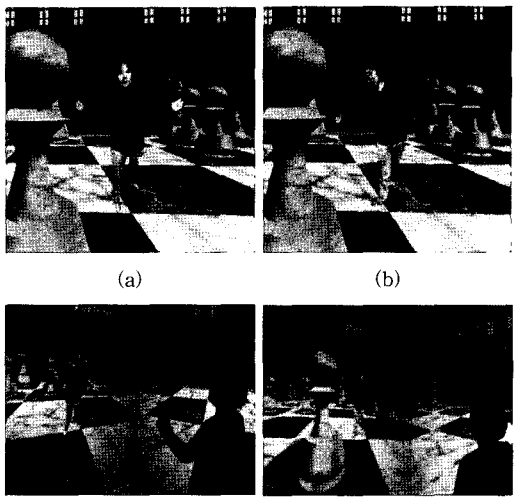
그림 8은 제안하는 방법에 따라 CAVE™-like 시스템 내에 있는 사용자 비디오 아바타를 생성하고 이를 가상세계에 투영한 예를 보여준다. (a)(b)(c)는 고화질 텍스처 결과를 제시하기 위하여 모노(mono) 디스플레이 된 결과영상이며, (d)는 스테레오 영상을 투영한 결과이다.

표 1 파이프라인 별 수행 시간 (단위 : sec)

	640×480	320×240
동기화 영상획득	0.015	0.014
실루엣 마스크 생성	0.059	0.013
텍스처-실루엣 피팅	0.047	0.012
압축 및 전송	0.055	0.001
총 계	0.176	0.040

6. 결론 및 향후 연구

본 논문에서는 어둡고 조명 조건이 일정하지 않은 CAVE™-like 시스템에서 시각 커뮤니케이션을 증진시켜주기 위한 고화질의 비디오 아바타 생성 기법에 대하여 소개하였다. 스크린으로부터 투사되는 빛의 변화를 차단하기 위하여 적외선 반사영상을 이용한 사용자 추출을 제안하였으며, 정의된 실루엣 마스크를 이용하여 텍스처 영상 내 배경을 제거하고 양안용 비디오 아바타 스테레오 영상을 생성하였다. 제안하는 방법은 기존의 방법과 비교했을 때 칼라카메라로 입력된 원영상을 그대로 이용하기 때문에 데이타 손실 없는 고화질의 결과영상을 생성할 수있다. 이 때 사람의 양안차 간격으로 배치된 카메라를 이용하여 양안용 영상을 각각 생성하고 프레임 순차 스테레오 방식을 적용하여 입체디스플레이 하기 때문에 사용자의 부피감을 그대로 재현할 수 있다. 따라서 삼차원 기하 정보의 재구성 없는 영상기반 방식임에도 불구하고 2.5D나 3D 비디오 아바타에서 제공되던 입체감을 살려줄 수 있으면서도 수행 속도가 빠르다는 강점을 지닌다.



(a) (b) (c) (d)

그림 8 스테레오 비디오 아바타 생성 및 투영 예: (a) (b) 모노 모드의 스크린 샷 (c) CAVE™-like 시스템 내에서의 모노 디스플레이 결과 (d) CAVE™-like 시스템 내에서의 스테레오 디스플레이 결과

향후 연구로는, 사용자 영상을 획득하는 카메라 위치에 따른 왜곡(foreshortening)을 보정하여 보다 사실적인 비디오 아바타 영상을 생성할 수 있도록 할 예정이며, 스크린 밝기에 따른 비디오 아바타 밝기를 정규화시키는 과정을 포함시킬 예정이다. 최종적으로는 국내 다른 연구기관에 설치되어 있는 CAVE™-like 시스템에 제안 방식을 적용하여 실제로 공유가상환경에서 비디오 아바타를 이용한 시각 커뮤니케이션의 유용성을 검증할 예정이다.

참 고 문 헌

[1] T. Ogi, T. Yamada, Y. Kurita, Y. Hattori, M. Hirose, "Usage of Video Avatar Technology for Immersive Communication," First International Workshop on Language Understanding and Agents for RealWorld Interaction, 2003.

[2] 김익재, 이상엽, 안상철, 권용무, 김형곤, "가상환경에서 다수의 실시간 비디오 아바타 생성기법", HCI 2003, 2003.

[3] 이원우, 우운택, "Network 가상환경을 위한 단순화 된 2.5D 비디오 아바타 생성", HCI2004, 2004.
[4] Youngjung Suh, Dongpyo Hong, Woontack Woo, "2.5D Video Avatar Augmentation for VR Photo," ICAT 2003, 2003.
[5] T. Ogi, T. Yamada, K. Tamagawa, M. Kano, M. Hirose, "Immersive Telecommunication Using Stereo Video Avatar," IEEE VR2001, 2001.
[6] V. Rajan, S. Subramanian, D.Keenan, A. Johswon, D. Sandin, T. Defanti, "A Realistic Video Avatar System for Networked Virtual Environments," IPT 2002, 2002.
[7] K.M. Cheung, S. Baker, and T. Kanade, "Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 2003.
[8] M. Gross, S. Wuemlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E.Koller-Meier, T. Svoboda, L. V. Gool, S. Lang, K. Strehlke, M. A. Vande, O. Staadt, "blue-c: A Spatially Immersive Display and 3D Video Portal for Telepresence," in Proceedings of ACM SIGGRAPH 2003, pp. 819-827, 2003.
[9] P. Debevec, C. Tchou, A. Wenger, T. Hawkins, A. Gardner, B. Emerson and A. Panday, "A Lighting Reproduction Approach to Live-Action Compositing," In Proceedings of the ACM SIGGRAPH 2002, 2002.
[10] Yasuda, K.,Naemura, T.,Harashima, H., 'Thermo-Key: Human Region Segmentation from Video, IEEE Computer Graphics and Applications," Vol. 24, No. 1, pp. 26-30, 2004.
[11] Sang-Yup Lee, Ig-Jae Kim, Sang C Ahn, Heedong Ko, Myo-Taeg Lim, Hyoung-Gon Kim, "Real Time 3D Avatar for Interactive Mixed Reality," In Proceedings of the ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI '04), Singapore, 16-18 June 2004.
[12] 박지영, 이선민, 김명희, "혼합현실환경을 위한 능동적외선 기반 동적배경 제거", 2004 한국정보과학회 추계학술발표대회 논문집, 2004.
[13] Matusik, Wojciech, "Image-Based Visual Hulls," Master of Science Thesis, Massachusetts Institute of Technology, 2001.
[14] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," In proceedings of IEEE Computer Vision and Pattern Recognition, 1986.
[15] A. Fusiello et al., "A Compact algorithm for rectification of stereo pairs," Machine Vision and Application vol.12 pp.16-22, 2000.
[16] S. Savarese, Camera Model and Triangulation, "Notes for EE-148 : 3D Photography," 2001.
[17] Rudolph E. Kalman, "An Introduction to Kalman Filter," University of North Carolina at Chapel Hill, Department of Computer Science, TR 95-041, 1995.
[18] M. Naef, O. Staadt, and M. Gross. blue-c api: A multimedia and 3d video enhanced toolkit for collaborative vr and telepresence. In Proc. ACM SIGGRAPH Int'l Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI'04), 2004.

이 선 민
1999년 이화여자대학교 컴퓨터학과 졸업(학사). 2001년 이화여자대학교 대학원 컴퓨터학과(공학석사). 2001년~현재 이화여자대학교 대학원 컴퓨터학과 박사과정. 관심분야는 가상증강 현실, 사용자 인터랙션, 영상가시화 등

박 지 영
2002년 이화여자대학교 컴퓨터학과 졸업(학사). 2004년 이화여자대학교 대학원 컴퓨터학과(공학석사). 2004년~현재 이화여자대학교 대학원 컴퓨터학과 박사과정. 관심분야는 컴퓨터 그래픽스, 가상현실 등

김 명 희
1979년 서울대학교 계산통계학과(석사) 1986년 독일 괴팅겐대학교 전자계산학과(박사). 1987년~현재 이화여자대학교 컴퓨터학과 교수. 1998년 7월~현재 이화여자대학교 컴퓨터 그래픽스/가상현실 연구센터장. 관심분야는 영상가시화, 시뮬레이션 및 가상현실 등