투영 텍스처를 이용한 스테레오 비디오 아바타+

이선민[1*] 박지영[1,2] 김명희[1,2]
[1]이화여자대학교 컴퓨터학과
[2]이화여자대학교 컴퓨터 그래픽스 / 가상현실 연구센터
(blue, lemie)@ewhain.net, mhkim@ewha.ac.kr

A Stereo Video Avatar using Projective Texture

Seon-Min Rhee[1*] Jiyoung Park[1] Myoung-Hee Kim[1,2]
[1] Dept. of Computer Science & Engineering, Ewha Womans University
[2] Center for Computer Graphics & Virtual Reality, Ewha Womans University

요 약

본 논문에서는 대형 스크린 기반 디스플레이 환경에서 시각 커뮤니케이션 증진을 위한 비디오 아바타 생성 기법에 대하여 소개한다. 일반적으로 대형 스크린 기반 가상 환경에서는 스크린으로 투사되는 빛의 변화 때문에 정적 배경을 보장할 수 없다. 뿐만 아니라 카메라 위치 설정이 자유롭지 못하기 때문에 디스플레이 환경 내에 있는 사용자를 대상으로 비디오 아바타를 생성하기 쉽지 않다. 본 논문에서는 적외선 및 필터를 이용하여 가시광선을 차단함으로써 사용자 영역을 쉽게 정의할 수 있도록 하였고, 나란히 배치된 컬러 영상에 실루엣 마스크를 적합하여 배경이 제거된 사용자 컬러 영상을 추출할 수 있도록 하였다. 또한 생성된 스테레오 비디오 아바타 영상은 투영 텍스처를 이용하여 보정된 후 가상세계에 투사되어 입체 디스플레이 된다. 제안된 기법은 일반적인 카메라를 이용하고 있으며, 간단한 조명 조건만 설정해주면 기존의 대형 스크린 기반 프로젝션 환경을 변형하지 않고도 쉽게 설치하여 사용할 수 있기 때문에 많은 활용도가 있을 것이라 기대된다.

1. 서론

비디오 아바타(video avatar)는 카메라로 획득한 실사 기반 사용자 모습을 가상세계에 투사하여 가상객체와 함께 보여주는 기법으로 원격실재감(tele-presence) 제공에 널리 이용된다. 특히, 가상협업환경과 같이 다른 가상환경내에 있는 상대방과의 원활한 의사 소통이 중요시 되는 경우, 시각 기반 커뮤니케이션(visual communication)을 증진시키기 위한 수단으로도 유용하게 활용될 수 있다. 최근에는 CAVE™-like 시스템과 같이 대형 디스플레이 환경을 네트워크로 연동하여 가상협업 환경을 구축하려는 시도가 늘고 있다. 이러한 환경에서는 비디오 아바타 생성 시, 스크린으로 투사되는 빛이 계속하여 변하기 때문에 배경 제거 (background subtraction)에 필수적인 정적 배경을 보장할 수 없으며 획득된 영상으로부터 사용자를 강건하게 추출하기 어렵다는 문제가 있다. 뿐만 아니라, 가상현실의 주요 요소인 사용자의 몰입감 (immersion)을 저해하지 않도록 하기 위해서는 카메라가 스크린을 가리지 않도록 해야 하는데, 이 경우 사용자 영상 획득에 최적화 된 카메라 위치를 보장할 수 없게 된다. 이와 같은 문제들로 인하여 대부분의 비디오 아바타 관련 연구에서는 사용자 영상을 획득하는 공간과 프로젝션 및 디스플레이 환경을 분리하고 있다. 그러나 이 경우에는 영상 획득 공간상에 있는 사용자는 자신의 모습을 상대방에 보여줄 수는 있지만 디스플레이 되는 상대방과 공유되는 가상세계를 볼 수는 없게 된다.

비디오 아바타를 생성하기 위한 기존의 연구들은 대부분 다수의 카메라로 입력 받은 영상으로부터 사용자의 형태를 삼차원으로 재구성하고 텍스처 매핑하여 가시화하는 방향으로 초점이 맞춰져 왔다. 한국과학기술 연구원(KIST)에서는 사용자 영상을 촬영하는 카메라와는 별도로 보조 카메라를 설치하여 사용자의 움직임을 추적함으로써 비디오 아바타가 보다 자연스럽게 가상공간 상의 다른 객체들과 합성될 수 있는 방법을 제안하였으며[1], 광주과학기술원(KJIST) 에서는 단순화 된 2.5D 비디오 아바타 생성 기법을 제안하였다[2,3]. T. Ogi et al.은 몰입형 가상 환경에서의 커뮤니케이션 수단으로써 비디오 아바타를 제안하고 있다[4]. 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

+ 본 논문은 정보통신부 대학정보통신연구센터(ITRC) 육성지원사업 수행 결과임

○ 카메라 캘리브레이션 및 렉티피케이션

흑백으로 정의된 실루엣 마스크를 이용하여 칼라 영상 내 배경을 제거하고 사용자 영역만 추출하기 위해서는 흑백 및 칼라 카메라 정보를 이용하여 두 영상 간의 상관관계를 사전에 파악해야 한다. 이를 위하여 전처리 단계로 Tsai가 제안한 카메라 캘리브레이션 방법[10]을 이용하여 각 카메라의 위치, 방향 정보를 획득하였다. 이는 실루엣 피팅 단계뿐 아니라 투영 텍스처 처리시 투영기의 위치 및 방향 설정에 이용된다. 또한, 두 영상의 카메라 중심을 나란하게 배치시킴으로써 실루엣 피팅 시 필요한 대응점 검색 차원을 낮추기 위해 필요한 렉티피케이션[11] 호모그래피 (homography) H_s, H_c를 각각 추출한다.

○ 실루엣 마스크 생성

가시광선이 차단된 흑백 영상에서 사용자 실루엣 마스크를 생성하기 위해서는 배경습득, 배경 분리, 후처리 단계를 거친다. 배경 습득 단계에서는 사용자가 존재하지 않는 배경 영상의 n개 연속 프레임을 이용하여 각 픽셀 위치의 m개 값에 대한 평균, 표준편차가 계산된다. 배경 분리 단계에서는 사용자가 존재하는 연속 영상에서 각 픽셀값과 배경습득 단계에서 계산된 같은 위치에서의 통계값을 식 (1)을 이용하여 비교함으로써 변화 여부를 판별한다. 식 (1)에서 p는 해당 픽셀, v는 픽셀값, μ 와 σ 는 각각 습득영상에 대해 사전에 계산된 각 픽셀 위치의 평균 및 표준 편차이고 k는 상수이다.

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

후처리 단계에서는 팽창(dilation)과 미디언 필터(median filter)를 적용하여 배경 분리 단계를 거친 영상의 잡음을 제거하고 사용자 영역을 보다 자연스럽게 나타나게 한다.

○ 실루엣 피팅

생성된 실루엣 마스크를 컬러 영상에 피팅하기 위한 단계는 다음과 같다. 우선 두 장의 실루엣 마스크 상의 중심점(centroid)을 대응점으로 가정하고 캘리브레이션을 통해 획득된 카메라 정보를 이용하여 삼차원 공간상의 사용자의 위치를 계산한다[12]. 삼차원 공간상의 위치를 계산하면 식 (2)를 이용하여 양안용 실루엣 마스크 및 칼라 영상간의 디스패러티를 각각 구할 수 있다.

$$disparity = \frac{b * f}{d} \quad (2)$$

(b :베이스라인 길이, f: 초점거리, d : 깊이값)

실루엣 피팅을 수행하기 위해서는 컬러 영상과 실루엣 마스크 영상간의 대응점을 찾아주어야 하며 이는 식(3)을 이용하여 계산할 수 있다. 이 때, $P_c(x_c, y_c)$는 컬러 영상내 화소 위치이며, $P_s(x_s,y_s)$는 실루엣 마스크 영상내 화소 위치이다. H_c, H_s는 전처리 단계에서 추출한 렉티피케이션 호모그래피이며, T는 식(2)에서 계산된 디스패러티 만큼의 수평 이동 행렬이다.

$$(x_s, y_s, 1)^T = H_c^{-1} * T * H_s * (x_c, y_c, 1)^T \quad (3)$$

식 (3)을 이용하여 컬러 영상 내에 각 화소에 대응되는 실루엣 마스크 영상에서의 위치를 찾고, 만약 찾아진 대응점이 실루엣 마스크 영역 내에 존재하면 해당 화소는 그대로 두며, 마스크 영역 외부에 존재하면 해당 화소 값은 사전에 정해진 배경색으로 수정한다. 따라서 그림 3에서 보는 것과 같이 실루엣 피팅을 수행하고 나면 사용자 영역만 남기고 배경을 삭제할 수 있게 된다. 이와 같은 단계를 거쳐 매 프레임 별로 양안에 해당하는 스테레오 사용자 영상을 생성한다.

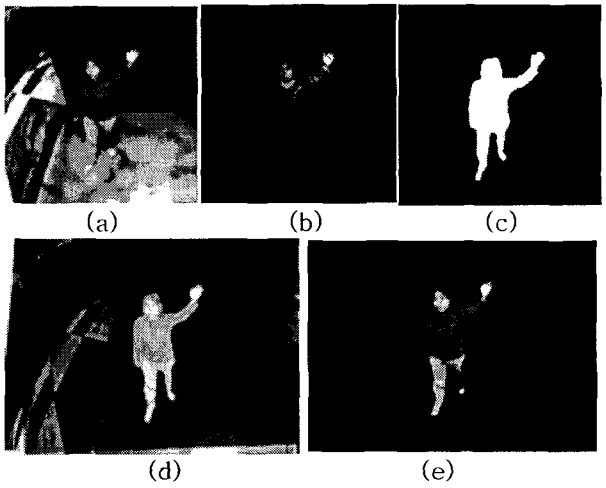


그림 3. 스테레오 비디오 아바타 생성을 위한 실루엣 피팅: (a) 컬러 영상 (b) 적외선 반사 흑백 영상 (c) 실루엣 마스크 영상 (d) 실루엣 피팅 (e) 배경 제거된 컬러 영상

2.3 가상 세계 내 비디오 아바타 투사

생성된 스테레오 영상은 압축되어 렌더링부로 전송되며 가상세계 내에 존재하는 빌보드에 텍스처 매핑하여 보여지게 된다. 이 때 사전에 배경색으로 지정된 화소는 투명 처리한다. 전송 시 해당 영상과 양안 정보(왼쪽 혹은 오른쪽)가 함께 보내지기 때문에 렌더링부에서는 동기화된 스테레오 영상을 제공할 수 있다. 따라서, 삼차원 정보를 재구성하지 않고도 사용자에게 입체감을 제공해 줄 수 있게 된다.

3. 투영 텍스처를 이용한 영상 보정

전면 상단에 설치된 카메라로부터 획득된 영상에서는 피사체 비율이 실제와는 다르게 나타난다(그림 4 참조). 즉, 본 실험 환경에서 획득된 사용자 영상을 그대로 비디오 아바타를 위한 텍스처로 사용하게 될 경우, 사람의 상체는 길고 하체는 짧게 나타나게 된다. 본 연구에서는 투영 텍스처(projective texture)[13]를 이용하여 영상

방법을 선택할 수 있는 근거를 제시하였다. V.Rajan et. al 은 네트워크 가상환경에서 사용자의 머리 모델을 삼차원으로 재구성하기 위한 방법을 제안한 바 있다[5]. 이와 같이 실사 기반 사용자 영상을 추출하여 비디오 아바타를 생성하고 이를 가상세계에 미러링 하기 위한 연구가 다양하게 진행되고 있지만, 대부분의 연구는 크로마키잉과 같이 배경 정적임을 전제하기 때문에 CAVE™-like 시스템과 같은 가상환경 내에 있는 사용자를 직접 촬영하지는 않고 있다.

프로젝션과 영상획득을 동시에 하기 위한 연구는 M.Gross et al.[6], P. Debevec et al.[7], Yasuda, K et al.[8], S.Y. Lee et al.[9]등에 의해 진행되어 왔다. 그러나 대부분의 연구는 특정 하드웨어를 사용하거나 두 카메라 사이의 빔 스플리터(beam splitter)나 반투명 거울(semi-transparent mirror)을 설치하여 빛의 일부는 투과시키고, 일부는 반사시킴으로써 두 카메라에 같은 영상을 동시에 입력시키는 방식을 이용한다. 특정 하드웨어를 사용하는 경우 기존에 설치되어 있는 가상환경에 적용될 수 없으며, 두 카메라를 동시에 이용하는 방법은 카메라의 위치 조정이 까다롭고, 빛의 일부가 손실되어 카메라로 입력되기 때문에 프로젝션 기반 가상환경이 어둡다는 점을 감안할 때 효과적이지 못하다

이와 같이 별도의 하드웨어 변형 없이 프로젝션 및 영상 획득을 동시에 가능하게 해주는 가상환경에 대한 연구는 드문 편이며, 배경색에 관계 없이 가상환경 내에 존재하는 사용자를 가상세계에 직접 미러링 하여 보여주는 예는 거의 없는 실정이다.

본 논문에서는 CAVE™-like 시스템과 같은 대형 스크린 기반 가상협업환경 내에 있는 사용자의 모습을 상대방에게 실시간 전송하여 시각 커뮤니케이션을 증진시키기 위한 수단으로 투영 텍스처를 이용한 스테레오 비디오 아바타 생성 기법에 대하여 소개한다. 스크린 상에 투사되는 빛을 차단하기 위하여 가시광선 차단 필터를 부착한 흑백 영상으로 사용자 실루엣 마스크를 정의하고 이를 칼라 영상에 적합하여 배경을 제거하였다. 또한 투영 텍스처를 이용하여 전면 상단에 설치한 카메라에 의해 획득된 영상의 왜곡을 보정하였다. 제안하는 방법은 별도의 하드웨어 변형 없이 기존의 디스플레이 장비에 쉽게 설치하여 이용할 수 있기 때문에 이미 대형 스크린 기반 가상환경을 가지고 있는 다양한 기관에서 쉽게 활용할 수 있을 것이라 기대한다.

본 논문의 구성은 다음과 같다. 2장에서는 CAVE™-like 시스템에서 비디오 아바타를 생성하기 위한 하드웨어와 시스템 개요에 대하여 간략하게 설명한다. 3장에서는 투영 텍스처를 이용한 영상 왜곡 보정에 관하여 기술하고 4장에서는 구현 및 결과를 보여준다. 5장에서는 결론 및 향후 연구 방향에 대하여 기술한다.

2. 시스템 개요

2.1 하드웨어 구성

본 연구에서는 실험 환경으로 전면, 좌측면, 우측면, 바닥면 네 개의 스크린으로 이루어진 대형 디스플레이 장비를 이용하였다. 사용자 영상 획득 시 필요한 카메라는 전면 상단에 설치하여 스크린을 가리지 않도록 함으로써 몰입감 저해를 최소화하였다. 스테레오 영상을 획득하기 위하여 칼라 카메라를 사람의 평균 양안차(약 65mm) 간격에 맞추어 배치하였다. (현재, 카메라 크기 때문에 80mm 간격으로 배치하였으나 이는 추후 소형 카메라를 이용함으로써 실제 양안차에 맞춰 최적화 시킬 수 있다.) 획득 영상으로부터 사용자를 강건하게 추출하기 위해서는 정적 배경을 보장해 주어야 한다. 이를 위하여, 카메라 렌즈에 가시광선 차단 필터를 부착하여 스크린 상에 투사되는 빛의 변화를 차단하였으며, 적외선 광원을 설치하여 적외선에 반사된 사용자가 카메라에 의해 감지될 수 있도록 하였다. 또한, 적외선을 방출하지 않는 주변 광원(ambient light)을 카메라 근처에 설치하여 사용자 영상 획득에 필요한 충분한 빛을 제공할 수 있도록 하였다. 실험 환경 내 카메라, 적외선 광원 및 주변 광원 위치 설정은 그림 1과 같다.

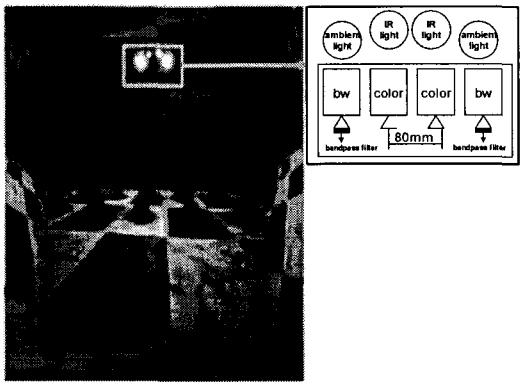


그림 1. 실험 환경 내 카메라 및 적외선 광원 위치

2.2. 스테레오 비디오 아바타 생성

비디오 아바타 스테레오 영상을 생성하고 이를 가상세계에 투사하여 가상객체와 함께 디스플레이 하기 위한 시스템 개요는 그림 2와 같다. 크게 영상 획득 및 처리부와 렌더링부로 나누어 볼 수 있으며 각 단계별 처리 내용은 다음과 같다.

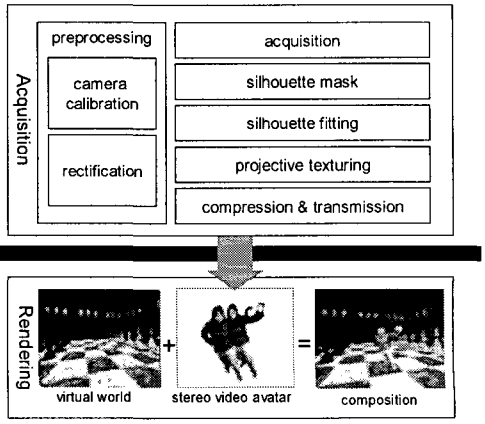


그림 2. 비디오 아바타 스테레오 영상 생성 파이프라인

내 피사체 비율을 재조정하였다.

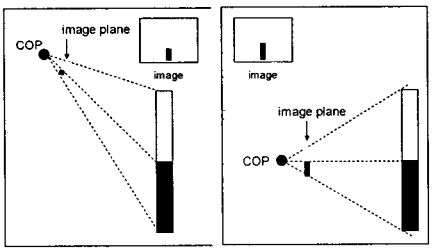


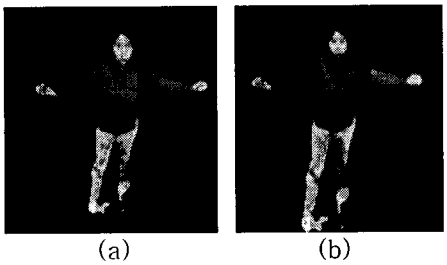
그림 4. 카메라 위치에 따른 영상에서의 피사체 비율 왜곡 비교: (a) 전면 상단부에 설치된 카메라 획득 영상 (b) 정면 중앙에 설치된 카메라 획득 영상

투영 텍스처에서는 투영기의 위치 및 방향을 사용자 임의로 조정할 수 있다. 따라서 이를 영상이 획득된 환경에서의 실제 카메라와 사용자간의 위치 및 방향과 동일하게 설정해주면 사용자의 신체 비율을 실제와 유사하게 복원할 수 있게 된다. 그림 5는 투영기 위치 및 방향 설정을 위한 OpenGL 코드이다.

```
glMatrixMode(GL_TEXTURE);
glLoadIdentity( );
glTranslatef(0.5, 0.5, 0.0);
glScalef(0.5, 0.5, 1.0);
glFrustum(xmin, xmax, ymin, ymax, focal, zmax);
glRotatef(xrot, 1.0, 0.0, 0.0);
glRotatef(yrot, 0.0, 1.0, 0.0);
glRotatef(zrot, 0.0, 0.0, 1.0);
glTranslatef(xpos, ypos, zpos);
```

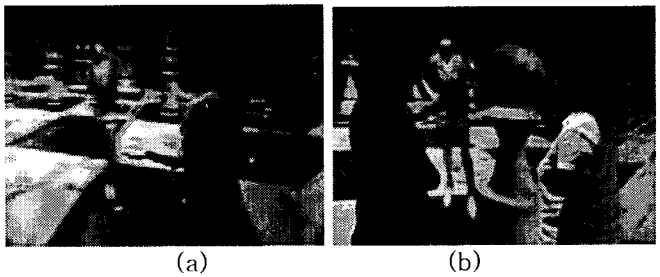
그림 5. 투영기 위치 및 방향 설정을 위한 OpenGL 코드: R(xrot, yrot, zrot)은 캘리브레이션을 통해 추출된 카메라 회전값, P(xpos, ypos, zpoz)는 실제 카메라와 사용자간 거리

그림 6은 투영 텍스처를 적용하기 전과 후의 결과를 비교한 예이다. 투영 텍스처 적용 전에는 다리 부분이 상대적으로 짧게 나타나지만, 투영 텍스처를 적용시키고 난 후에는 사용자의 실제 신체 비율로 보정됨을 알 수 있다.



(a) (b)

그림 6. 투영텍스처 (a) 적용 전 및 (b)적용 후 결과 비교



(a) (b)

그림 7. CAVE™-like 시스템에서의 비디오 아바타 (a) 사용자 미러링 (b) 비디오 아바타를 이용한 협업

4. 실험 결과

실험에 사용된 컴퓨터는 Dell Dimension 8300(Intel Pentium Pentium 4 Dual CPU 3.00GHz, 2GB RAM) 이며, 카메라 네 대를 하나의 컴퓨터에 연결하여 카메라 간 동기를 쉽게 맞출 수 있도록 하였다. 사용된 카메라는 Point Grey Research Inc. 의 Dragonfly 이며, 최대 해상도는 640x480, 최대 프레임 레이트는 30fps 이다. 실험에 사용된 영상은 640x480, 320x240, 15fps 이다. 영상 처리에 필요한 라이브러리는 Intel 사의 OpenCV를 이용하였으며, 렌더링부에서는 bluc-c API[14]를 이용하여 가상세계를 모델링하고, 스테레오 비디오 아바타를 텍스처 매핑하여 디스플레이 하였다. 각 단계별 수행 시간은 표 1과 같다.

표 1. 각 스테레오 비디오 아바타 생성 단계별 수행 속도

단계 \ 해상도	640x480	320x240
영상 획득	0.015	0.014
배경 제거	0.059	0.013
실루엣 피팅	0.047	0.012
텍스처 투영	0.047	0.011
합	0.168	0.050

640x480의 해상도일 경우 스테레오 영상 생성시 소요 시간은 0.168초로 약 6fps이며, 320x240의 경우 0.05초로 20fps 이다. 따라서, 여러 대의 컴퓨터를 이용하고, 각 단계를 쓰레드 화 하여 파이프라이닝으로 처리 하는 등의 최적화 작업을 통하여 실시간에 가까운 속도로 고화질의 비디오 아바타 스테레오 영상을 생성할 수 있을 것이라 기대할 수 있다. 그림 7(a)는 CAVE™-like 시스템 내에 있는 사용자의 비디오 아바타를 생성하여 가상세계에 미러링 한 결과 영상이다. 그림 7(b)는 비디오 아바타를 이용하여 상대방 가상환경 내에 있는 사용자와 협업하고 있는 결과 영상이다.

5. 결론 및 향후 연구

본 논문에서는 대형 디스플레이 환경에서 텔레프레센스

및 시각 기반 커뮤니케이션 지원 수단으로 활용할 수 있는 스테레오 비디오 아바타 생성 기법에 대하여 소개하였다. 제안 기법에서는 스크린 상에 투사되는 빛의 변화를 차단하고 강건하게 실루엣을 추출하기 위하여 적외선을 이용한 실루엣 마스크를 생성하였고 이를 이용하여 칼라 영상 내 사용자를 둘러싼 배경을 쉽게 제거할 수 있도록 하였다. 또한 투영 텍스처 방식을 이용하여 카메라 위치에 따른 피사체 비율을 보정을 시도하였다. 이 방식은 삼차원 모델을 이용한 비디오 아바타 생성시 발생하는 컬러 정보의 손실이 거의 일어나지 않기 때문에 보다 고화질의 결과를 보장할 수 있다. 또한, 특별한 하드웨어 변형 없이 일반 카메라를 이용하며 간단한 조명 조건 설정만으로 쉽게 적용할 수 있기 때문에 기존의 대형 VR 장비를 가지고 있는 연구소나 대학에서 쉽게 기술을 활용할 수 있을 것이라 기대한다.

향후 연구로는 단일 사용자 뿐 아니라 다수 사용자 비디오 아바타를 생성할 수 있도록 확장하여 그룹간의 협업을 위한 증진된 시각 기반 커뮤니케이션을 지원할 수 있도록 할 예정이다.

참고문헌

[1] 김익재, 이상엽, 안상철, 권용무, 김형곤, 가상환경에서 다수의 실시간 비디오 아바타 생성기법, *HCI 2003*, 2003

[2] 이원우, 우운택, Network 가상환경을 위한 단순화 된 2.5D 비디오 아바타 생성, *HCI2004*, 2004.

[3] Youngjung Suh, Dongpyo Hong, Woontack Woo, 2.5D Video Avatar Augmentation for VR Photo, *ICAT 2003*, 2003.

[4] T. Ogi, T. Yamada, K. Tamagawa, M. Kano, M. Hirose, Immersive Telecommunication Using Stereo Video Avatar, *IEEE Virtual Reality 2001*, 2001

[5] V. Rajan, S. Subramanian, D.Keenan, A. Johswon, D. Sandin, T. Defanti, A Realistic Video Avatar System for Networked Virtual Environments, *IPT2002*, 2002

[6] 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, *ACM SIGGRAPH 2003*, pp. 819-827, 2003.

[7]vP. Debevec, C. Tchou, A. Wenger, T. Hawkins, A. Gardner, B. Emerson and A. Panday, A Lighting Reproduction Approach to Live-Action Compositing, *ACM SIGGRAPH 2002*, 2002

[8] Yasuda, K., Naemura, T., Harashima, H., Thermo-Key: Human Region Segmentation from Video, *IEEE Computer Graphics and Applications*, 24(1): 26-30, 2004.

[9] 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, *ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI'04)*, Singapore, 16-18 June 2004.

[10] R. Y. Tsai, An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, *IEEE Computer Vision and Pattern Recognition*, 1986

[11] A. Fusiello et al., A Compact algorithm for rectification of stereo pairs, *Machine Vision and Application* vol.12:16-22, 2000

[12] S. Savarese, Camera Model and Triangulation, *Notes for EE-148 : 3D Photography*, 2001.

[13] Mark Segal, et al. Fast shadows and lighting effects using texture mapping. *In Proceedings of SIGGRAPH ? 2*, pages 249-252, 1992.

[14] M. Naef, O. Staadt, M. Gross : Multimedia integration into the blue-c API.*Journal of Computers & Graphics*, vol. 29, issue 1, pp. 3-15, February 2005