

# 실제와 합성영상의 비교에 의한 비디오 아바타의 정합

## (Registration of Video Avatar by Comparing Real and Synthetic Images)

박 문 호 <sup>†</sup>      고 희 동 <sup>\*\*</sup>      변 혜 란 <sup>\*\*\*</sup>  
 (Moon Ho Park)    (Heedong Ko)    (Hye Ran Byun)

**요 약** 본 논문에서는 가상 환경의 참여자를 표현하기 위해 실제 참여자의 영상을 가상 환경에 실시간으로 제공하는 비디오 아바타를 사용하였다. 비디오 아바타의 사용은 참여자를 표현하는 정밀도를 높일 수 있지만 정확한 정합이 중요한 이슈가 된다. 비디오 아바타의 정합을 위해 실제 환경에서 사용되는 카메라와 가상 환경을 생성하기 위해 사용되는 가상 카메라의 특성을 동일하게 조정하였다. 조정된 실제와 가상 카메라의 유사성에 근거하여 실제와 합성 영상의 비교를 통하여 실제 환경에서 획득된 비디오 아바타가 가상 환경과 정합되도록 하였다. 비디오 아바타의 정합 과정에서는 정합의 부정확한 정도를 에너지로 표현하여 이를 최소화시키는 방법을 이용하였으며 실험을 통하여 제안된 방법이 가상 환경에서 비디오 아바타의 정합에 효과적으로 적용 가능함을 확인하였다.

키워드 : 가상환경, 비디오 아바타, 정합, 혼합현실, 블루룸

**Abstract** In this paper, video avatar, made from live video streams captured from a real participant, was used to represent a virtual participant. By using video avatar to represent participants, the sense of reality for participants can be increased, but the correct registration is also an important issue. We configured the real and virtual cameras to have the same characteristics in order to register the video avatar. Comparing real and synthetic images, which is possible because of the similarities between real and virtual cameras, resolved registration between video avatar captured from real environment and virtual environment. The degree of incorrect registration was represented as energy, and the energy was then minimized to produce seamless registration. Experimental results show the proposed method can be used effectively for registration of video avatar.

**Key words** : Virtual Environment, Video Avatar, Registration, Mixed Reality, Blue Room

### 1. 서 론

가상환경을 이용한 여러 가지 적용 사례를 보면 가상 환경에 대한 사실적인 표현을 통하여 사용자들에게 가상 환경에 대한 신뢰감과 작업의 효율성을 높이고 있다 [1-2]. 가상환경을 사실적으로 표현하기 위하여 실제 환경의 정보를 가상 환경에 포함하여 가상 환경을 보완

또는 증강시키는 방법이 사용된다. 실제 환경에서 발생한 시각적 정보를 가상환경에 가시화를 통하여 실제 환경과 가상 환경이 서로 연결되는 것을 혼합현실(Mixed Reality) 또는 증강현실(Augmented Reality)이라고 한다[3,4]. 한편, 가상 환경에서 중요한 구성요소중 하나인 참여자를 다각형 모델로 표현하기도 하지만 단순하게 표현된 인체 데이터로 인하여 사실감이 감소하게 된다. 사실감을 증대시키기 위해 더 많은 다각형이 사용되며 구체적으로 표현하면 할수록 모델링과 렌더링 과정에서 추가적인 부담이 발생된다. 반면에 가상 환경에서 참여자를 표현하는데 있어서 현실 세계로부터 전달되는 비디오 영상을 사용하게 되면 참여자를 사실적으로 표현할 수 있는 장점이 있어서 실제로 방송 분야에서 활용되고 있다[5]. 그러나 현실 세계에서 얻어진 비디오를

· 이 논문은 한국학술진흥재단의 지원에 의하여 연구되었음(02-13-04-1)

† 정 회 원 : 삼성전자 연구원  
moonhopark@naver.com

\*\* 정 회 원 : 한국과학기술연구원 영상미디어연구센터 연구원  
ko@kist.re.kr

\*\*\* 정 회 원 : 연세대학교 컴퓨터과학과 교수  
hrbyun@cs.yonsei.ac.kr

논문접수 : 2005년 8월 12일

심사완료 : 2006년 4월 3일

가상 환경에서 사용하기 위해서는 비디오를 가상의 환경과 정확히 연결시키는 과정이 필요하다.

이러한 문제를 해결하기 위해 본 논문에서는 실제와 가상 카메라의 속성을 동일하게 하고 실제와 합성 영상의 비교를 통하여 실제 세계의 비디오가 가상 환경과 조화를 이루도록 하는 방법을 제안하였다. 본 논문은 책상 주변에서 참여자들이 상호 작용하는 환경을 대상으로 하며, 따라서 책상 주변에 앉아서 작업하는 참여자의 비디오 영상을 가상환경에 정합시키게 된다. 본 논문의 구성은 다음과 같다. 1장의 서론에 이어서 2장에서는 비디오 아바타에 관련된 연구를 소개하며, 3장과 4장에서 제안된 정합방법과 실험 결과를 각각 설명한다. 마지막으로 5장에서 결론 및 향후 연구 과제를 계속해서 설명한다.

## 2. 비디오 아바타 관련 연구

아바타는 가상공간에서 참여자의 역할을 하는 객체이며, 아바타를 통하여 가상공간 안에서 참여자들 사이에 의사소통을 하게 된다[6]. 아바타는 참여자의 위치나 보는 방향, 감정 등을 나타내게 되는데 실시간 비디오를 통하여 참여자의 모습을 구현하게 되면 자연스러운 얼굴표정이나 몸동작을 통하여 사실감을 높일 수 있는 효과를 거둘 수 있을 뿐만 아니라 다른 사람을 3차원 공간에서 아바타를 통하여 만나서 서로 인식하는 것도 가능하게 된다. 이와 같이 비디오 영상을 통하여 가상환경의 아바타를 표현하는 형태를 비디오 아바타라고 부른다[7,8].

비디오 아바타는 2차원 평면이나 복셀(voxel) 형태로 구현되며, 중간형태인 2.5차원의 형태로도 표현된다. 이러한 형태에 따라서 표현력의 차이가 발생하게 되며, 필요에 따라서는 얼굴을 중심으로 하는 인체의 일부분만을 이용하여 표현하기도 한다. 본 논문에서는 구현하기 용이하면서 참여자를 쉽게 표현할 수 있는 장점이 있는 평면 모델을 비디오 아바타로 사용하였다. 이때 사람을 제외한 다른 영역을 제거하게 되면 비록 비디오 아바타가 다각형에 비디오 매핑한 텍스처에 불과하지만 3차원 가상공간 안에서 몰입적이고 사실적인 형태로 설득력 있게 참여자를 표현할 수 있는 수단이 된다. 배경을 포함하고 있는 영상으로부터 참여자를 분할하기 위해서는 크로마키 방법을 이용하거나 배경 영상의 통계적 특성을 이용하여 참여자를 배경으로부터 분리하는 방법이 있다. 본 논문에서는 크로마키 기법을 사용하였고, 따라서 사용자가 블루(blue) 색상 영역의 배경 앞에 있어야 하고, 참여자 이외의 색상 영역을 투명하게 처리한 후 텍스처로 만들어 3차원 가상공간 안에 삽입하였다.

이제 3차원 공간에 삽입된 비디오 아바타가 사실적으

로 표현되기 위해서는 실제 환경의 영상과 가상환경의 요소를 동일한 좌표 공간에 존재하는 것처럼 조정하는 과정이 필요하게 된다. 증강현실 응용 프로그램을 구현하기 위해 많이 사용되는 ARToolkit에서는 비디오 아바타의 위치와 방향을 마커를 이용하여 결정하게 된다. 이 방법에서는 마커를 바라보는 카메라의 상대적인 값을 구하여 마커에 위치하게 되는 비디오 아바타의 형태를 결정하게 된다[9]. 양안시차에 따른 깊이 정보를 이용하여 비디오 아바타와 3차원 가상 객체를 정합하는 방법이 사용되기도 한다. 카메라의 속성정보를 이용하여 전경과 배경사이의 거리를 계산하여 정확한 Z-keying이 가능하게 된다[10]. 가상 스튜디오 관련 기술을 살펴보면 비디오 영상을 촬영하는 실제 카메라와 컴퓨터 그래픽으로 배경을 렌더링하는 가상 카메라와 연동을 통하여 비디오 영상을 가상세계에 정합하게 된다. 카메라에 부착된 기계적 또는 적외선 센서를 이용하거나 블루룸 안에 마련된 패턴을 이용하여 실제 카메라의 위치 정보를 얻게 되고, 이 정보를 가상 카메라와 연동시킴으로써 비디오 영상을 가상세계에 정합시키게 된다[11].

## 3. 제안된 비디오 아바타 정합 방법

혼합현실에서 현실 세계의 객체와 가상 세계 사이의 관계를 공간적 또는 시각적으로 일치시키는 것을 정합이라고 한다[3]. 사실감을 증대시키기 위하여 현실과 가상 세계의 객체들은 가상과 현실 세계 안에서 상호간 또는 제3자의 관점에서 정확히 맞추어져서, 두 개의 이질적인 구성 요소들이 같은 좌표 공간 안에 동시에 존재하고 있는 효과를 생성해야한다. 비디오 아바타를 정확한 장소에 위치시키기 위해서 수작업으로 비디오 아바타의 위치를 찾게 되면 정확한 값을 찾기 불가능하다는 문제가 있다. 본 논문에서는 비디오 아바타를 가상환경에 정합하는 과정을 자동화하기 위한 방법으로 카메라의 유사성을 근거로 하여 실제와 합성 영상을 비교하여 차이를 줄여가는 방법을 사용하였다.

### 3.1 카메라의 유사성

비디오 아바타를 표현하는 다각형의 위치와 방향 그리고 스케일 등을 정확히 조정해 주어서 주변의 가상환경과 조화를 이루게 하는 과정을 필요로 한다. 이러한 비디오 아바타의 파라미터를 구하기 위하여 본 논문에서는 실제 카메라와 가상 카메라의 특성을 이용하였다. 가상환경을 표현하는 컴퓨터 그래픽스 기법은 합성될 영상을 렌더링하기 위한 객체나 광원 또는 카메라의 속성 등의 자료를 이용하여 영상을 생성하는 과정을 거치게 된다. 따라서 가상세계를 기술하는 최적의 파라미터를 구하기 위하여 실제세계에서 카메라로 촬영한 영상과 컴퓨터 그래픽으로 생성된 영상과의 차이를 비교하

여 두 영상의 차이가 최소화되도록 합성된 영상을 기술하는 파라미터를 최적의 해로 구하는 방법을 이용하였다. 그림 1에서 이러한 과정을 수행하기 위한 과정을 나타내고 있다.

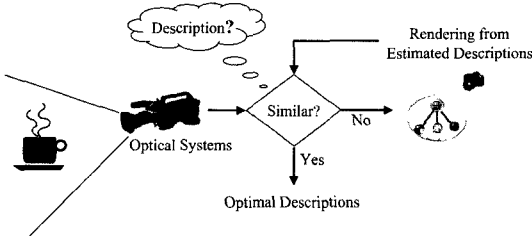


그림 1 반복수행에 의한 파라미터 추출과정

실제 카메라에서 캡처된 영상과 비디오 아바타가 렌더링되는 가상환경을 생성하는 가상 카메라에서 출력되는 영상과의 비교를 통하여 비디오 아바타의 최적의 파라미터를 구하게 된다. 즉, 실제 영상과 렌더링된 영상의 차이를 가장 최소화하는 파라미터를 찾는 것이다. 이를 위해 두 영상의 차이를 에너지로 나타내는 함수로 표현하여 에너지 최소화 기법을 사용하여 원하는 값을 구하게 된다[12]. 이러한 과정을 수행하기 위해 선행되어야 되는 절차가 실제 카메라와 가상 카메라가 동일한 특성을 가지도록 조정하는 것이다. 실제 카메라는 다음의 그림 2와 같은 특성을 가지고 있다.

물리적 카메라에서 영상의 생성과정은 그림 2에서 볼 수 있듯이 무한대에서부터 카메라의 렌즈 앞까지 광범위한 영역에 존재하는 무수히 많은 점  $m(X,Y,Z)$ 이 카메라의 렌즈를 통과하여 카메라의 필름에 도달하게 된다. CCD 카메라의 경우에는 필름 대신에 CCD 센서 배열에 3차원 공간에 존재하는 점  $m$ 이 도달하게 된다.

가상 카메라에서는 가상세계에 존재하는 점  $m(X,Y,Z)$ 이 원근 투시(Perspective Projection) 과정을 거쳐 시점(viewpoint) 앞에 있는 뷰 평면(view plane)에 도달하게 된다. 인간의 눈에 영상이 맺히는 경우와 물리적 카메라의 경우와는 다르게 가상 카메라에는 시야 범위의 제한이 있다. 즉, 뷰 평면에 투영되는 범위가 가상

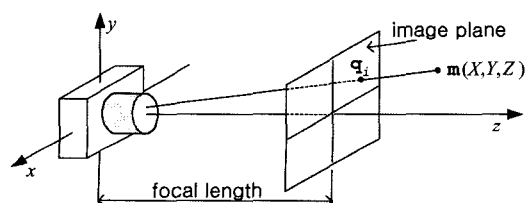


그림 2 실제 카메라 시스템

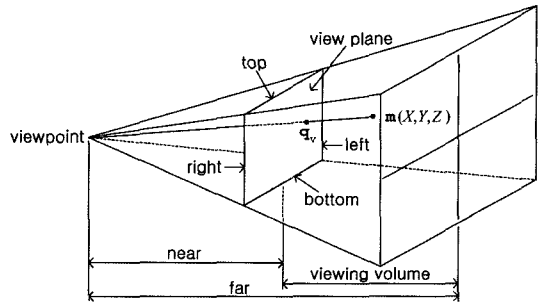


그림 3 가상 카메라 시스템

카메라에서는 카메라 파라미터인 near와 far로 제한되게 되며 상하좌우도 top, bottom, left, right로 엄격하게 제한 되게 된다. 그림 3에서처럼 뷰잉 볼륨(viewing volume)의 영역 안에 존재하는 물체만 볼 수 있게 되며 나머지 부분은 차단되어 진다.

### 3.2 가상과 실제의 시각적 환경 조정

지금까지 살펴본 바와 같이 물리적인 실제 카메라와 컴퓨터 그래픽스 시스템이 생성하는 가상의 카메라는 유사성을 가지고 있다. 따라서 이러한 유사성을 이용하여 3차원 가상환경에 존재하게 될 비디오 아바타의 위치나 방향을 결정하기 위하여 실제와 가상의 카메라의 특성이 같도록 카메라 파라미터를 조정하는 과정이 필요하다. 실제와 가상의 카메라 유사성은 다음의 식으로 비교된다.

$$q_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & f/0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P_o m$$

$$q_v = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & d/0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P_g m \quad (1)$$

여기서  $q_i = [x_i y_i z_i w]^T$ 와  $q_v = [x_v y_v z_v w]^T$ 는 각각, 실제 카메라의 영상 평면(image plane)과 가상 카메라의 뷰 평면의 한 점을 의미하며, 이 점은  $P_o$ 와  $P_g$ 로 표현되는 광학적 원근 투영 과정과 그래픽 원근 투영 과정을 거친 3차원 공간의 한 점  $m$ 이 투영된 결과로 발생한 점이다.  $f$ 와  $d$ 는 각각 실제 카메라의 초점 거리(focal length)와 시점과 뷰 평면의 중심과의 수직 거리를 의미한다.

가상 카메라의 뷰잉 볼륨은 그림 3에서 표시된 것처럼 left, right, bottom, top, near, far의 6개 파라미터로 결정되어진다. 따라서 실제 카메라의 내부 파라미터[13]를 이용하여 실제 카메라와 동일한 뷰잉 볼륨을 가진 가상 카메라를 설정할 수 있다. 즉, near를 초점 거리와 동일한 값으로 설정하고 left, right, bottom, top은 실제

카메라의 CCD 센서 배열의 물리적인 넓이와 높이를 이용하여 시점을 기준으로 상하좌우 대칭적으로 결정하며, 마지막으로 far는 충분히 큰 값으로 설정하여 가상 카메라가 바라보는 물체가 차단되지 않도록 한다. 이러한 과정으로 두 카메라의 특성을 같게 하면 두 카메라의 보는 환경이 동일하게 된다[14,15]. 카메라 캘리브레이션으로 얻은 외부 파라미터를 이용하여 실제 카메라의 위치에서 가상 카메라를 위치시킬 수 있다. 실제 카메라와 가상 카메라의 속성이 동일하게 되면 두 카메라로부터 얻어지는 내용도 동일하게 된다. 따라서 비디오 아바타를 렌더링한 결과와 실제 카메라에서 얻어진 영상과의 비교를 통해 비디오 아바타에 대한 최적의 파라미터를 추출할 수 있게 된다. 두 카메라가 유사하기는 하지만 불가피한 차이를 가지고 있다. 물리적 카메라의 렌즈 왜곡이나 가상 카메라의 거리에 따른 초점의 균일성 등으로 들 수 있다. 본 논문에서는 실제 카메라 영상과 가상 카메라의 영상을 비교하기 위하여, 실제 영상을 변형(warping)하여 렌즈 왜곡을 제거하였다.

**3.3 에너지 최소화 방법**

실제 카메라로 촬영한 영상과 비디오 아바타를 대략적으로 예측된 위치에 놓고 렌더링한 영상과 비교함으로써 최적의 비디오 아바타 파라미터를 찾기 위하여 LM(Levenberg-Marquardt) 방법을 사용한다[12]. LM 방법은 에러 함수의 2차 테일러 전개식을 기반으로 하는 최적화 방법이다. 이 반복적인 방법의 시작은 초기 파라미터의 추정치를 사용하며 매번 반복되는 과정에서 에러 함수를 줄이는 파라미터를 추출하게 된다. 추출된 파라미터는 다음 반복에 반영되며 종료 조건이 만족될 때까지 이러한 과정을 반복한다. 주어진 뷰잉 볼륨 안에서 비디오 아바타의 임의의 위치에서 기하학적인 정보와 색 정보를 이용하여 영상이 렌더링 된다. 한편, 최적의 비디오 아바타의 파라미터는 목표로 하는 영상과 가장 비슷한 영상을 렌더링 하는 비디오 아바타의 파라미터를 탐색함으로써 구해질 수 있다. 비디오 아바타의 파라미터는 비디오 아바타를 표현하는 평면의 3차원 가상 공간 안에서의 위치와 방향이다. 비디오 아바타의 스케일은 비디오가 매핑되는 평면 안의 실제 영상의 크기와 가상환경에서 해당하는 크기를 비교함으로써 구하게 된다. 실제 카메라로 촬영한 영상을  $\psi'$ 로 표현하면 에러 함수는 다음과 같이 표현된다.

$$E = \|\psi' - \psi(\hat{\alpha}|\mathbf{u})\|^2 \tag{2}$$

여기서  $\psi(\hat{\alpha}|\mathbf{u})$ 는 주어진 뷰잉 볼륨( $\mathbf{u}$ )안에서 추정된 비디오 아바타의 파라미터( $\hat{\alpha}$ )에 따라 렌더링된 영상을 의미한다. 그림 4는 반복적으로 수행되는 비디오 아바타

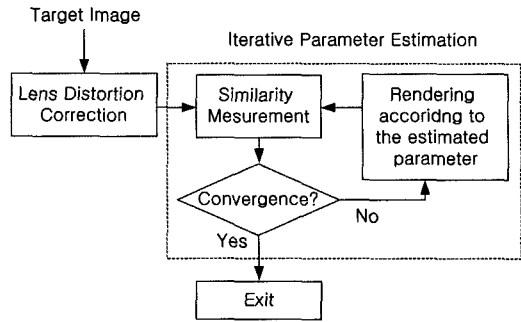


그림 4 비디오 아바타의 파라미터 추출과정

의 파라미터 추출 과정을 설명하고 있다.

각 반복 과정에서 구해진 파라미터의 변화량( $\Delta\hat{\alpha}$ )은 다음 반복 과정의 파라미터를 결정하는데 반영되며 파라미터의 변화량을 구하는 식을 표현하기 위해 Newton의 최소화 방법을 n차원의 경우로 확장하면 다음과 같이 표현된다.

$$W_{i+1} = W_i + H^{-1}(W_i)\nabla g(W_i) \tag{3}$$

$H^{-1}$ 는 Hessian 매트릭스의 역(inverse)을  $\nabla$ 은 그레디언트(gradient)를 의미한다. 한편,  $W_0$ 에서 Taylor의 전개식을 이용하여 풀면 다음과 같다.

$$E(W) = E(W_0) + \nabla E(W_0)(W - W_0)^T + 1/2(W - W_0)H(W_0)(W - W_0)^T + \dots \tag{4}$$

양변에  $\nabla$ 를 취하여 정리하면 식 (5)를 얻는다.

$$W = W_0 - H^{-1}(W_0)\nabla E(W_0) \tag{5}$$

한편, 기울기 강하(Steepest Descent) 방법은 식 (6)과 같이 표현된다.

$$W_{i+1} = W_i - \lambda \nabla E(W_i) \tag{6}$$

여기서,  $\lambda$ 는 학습률을 의미한다. Newton의 방법은 최소 값 근처에서 빠른 속도로 수렴하지만 최소 값과 최대 값을 동시에 탐색하기 때문에 안장점에 빠질 우려가 크다. 따라서 LM 방법은 Steepest Descent 방법과 Newton의 방법을 동적으로 조정하며 최적의 해를 찾는 방법을 채택하며 다음의 식으로 표현된다.

$$W_{i+1} = W_i - (H + \mu I)^{-1} \nabla E(W_i) \tag{7}$$

식 (7)에서 만약  $\mu$ 가 0이면 Newton의 방법이 수행되고,  $\mu$ 가 커지면 Steepest Descent 방법을 수행하게 된다. LM 방법에서는 Hessian 매트릭스를  $J^T J$ 로 근사화하게 되는데, 여기서  $J$ 는 자코비안 매트릭스(Jacobian Matrix)를 나타낸다. 따라서  $\nabla^2 F = 2J^T J$ 로 근사화되며 식 (8)로 표현 가능하다.

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)E(x_k) \tag{8}$$

식 (7)의  $\mu_k I$ 를  $J^T(x_k)J(x_k)$ 에 추가하면 식 (8)이 유

도된다. 여기서,  $\mu_k$ 는 k번째 단계에서 조정되는 파라미터를 의미한다. 본 논문에서는 자코비안을 근사화하기 위해 식 (9)와 같이 중앙 차분을 이용하였다.

$$dI/d\alpha_i(\hat{\alpha}) \approx (I(\alpha + \Delta\alpha_i e_i) - I(\alpha - \Delta\alpha_i e_i)) / 2\Delta\alpha_i \quad (9)$$

여기서,  $e_i$ 는 기저(basis)의 i번째 벡터이고,  $\Delta\alpha_i$ 는 파라미터  $\alpha_i$ 의 증분이다. 이러한 과정으로 유도된 식 (8)을 이용하여 각 반복과정에서 변화량을 계산하고 다음 단계에서의 파라미터를 추정하게 된다. 지금까지 기술한 에너지 최소화 방법의 수행 과정을 그림 5에 도시하였다. 이러한 과정에서 목표영상과 렌더링된 비디오 아바타의 차이를 나타내는 에러함수는 배경이 차단된 전경을 이진화한 영상에 마스크를 사용하여 특징 값을 추출하였고, 이러한 특징 값을 식 (10)과 같은 에러함수의 값으로 사용하였다.

$$E = \sum_{p=1}^{10} \sum_{q=1}^{10} \sum_{i=1}^n \sum_{j=1}^m I_{pq}(i, j) \quad (10)$$

여기서, p와 q는 각각 세로방향의 p번째와 가로방향

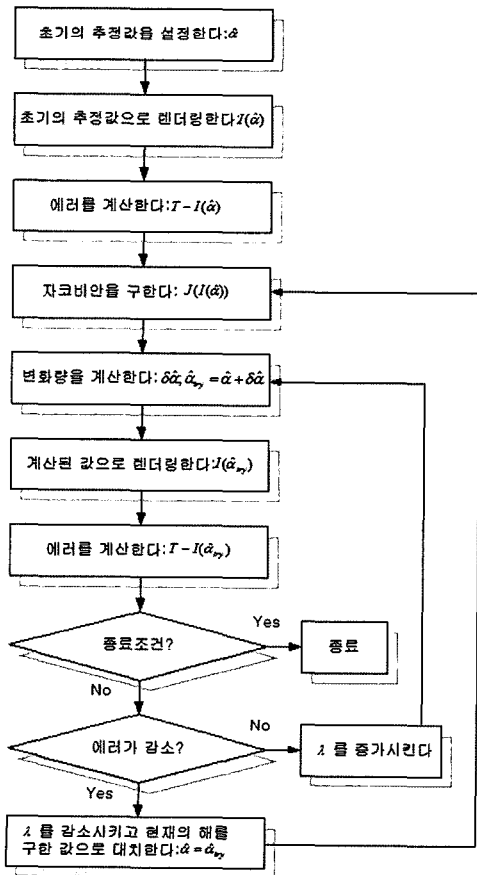


그림 5 에너지 최소화 방법의 수행 과정

으로 q번째의 블록을 의미하며,  $I_{pq}(i, j)$ 는 pq블록의 i와 j번째 위치에서의 특징 값을 나타낸다. 이 식에서 표시된 것처럼 가로와 세로를 10등분으로 균등하게 분할하여 총100개의 블록에서 특징을 추출하였다. 이러한 에러함수의 값이 적으면 적을수록 비교되는 두 영상은 유사한 형태를 가지게 되며 구하고자 하는 비디오 아바타의 최적의 파라미터를 추출할 수 있게 된다.

#### 4. 실험 결과

본 논문에서 구현된 시스템은 가상환경을 렌더링하기 위해 Silicon Graphics사의 Onyx I IR을 사용하였고 가상환경에 비디오를 연동하기 위하여 Sirius 비디오 보드를 이용하였다. 가상환경과 객체의 모델링은 Multi-Gen을 이용하였다. Onyx I에 의하여 렌더링 되는 가상환경은 사용자들에게는 몰입적인 디스플레이 장치인 HMD(Head Mounted Display)를 통하여 시각적인 정보로 전달하게 된다. HMD는 n-Vision 스테레오 HMD를 사용하였고, 사용자의 머리 움직임을 추적하여 실시간 렌더링에 반영하기 위해 전자기계 위치추적 장치인 Polhemus사의 Fastrack을 이용하였다. Fastrack은 전자기판을 송출하는 송신장치와 이를 수신하여 위치정보를 본체에 전달해 주는 수신장치로 구성되며 수신장치가 HMD에 부착되어 위치를 추적하게 된다. Fastrack의 정보는 Onyx I에 RS-232 시리얼 통신으로 전달되게 된다. 따라서 사용자가 머리를 움직이는 동작이 가상환경에 실시간으로 반영되어 사실감과 존재감을 높여주게 된다. 그림 6은 구현된 시스템의 전체적인 구성을 나타내고 있다.

블루 룸에서 캡처되는 실제 영상을 Matrox DigiSuite의 DigiMix를 이용하여 크로마키 처리를 한다[5]. 크로마키 처리된 영상은 알파 값을 이용하여 배경이 투명하게 처리되며, 가상환경 안에서 비디오 텍스처로 사용된다. 즉, 카메라로 촬영된 영상은 DigiMix로 입력되며 크로마키 작용을 거친 RGBA의 영상 신호를 Onyx I에 연결된 Sirius 비디오 보드에 입력한다. 시스템은 OpenGL Performer를 기반으로 가상 환경을 렌더링하며[16] Sirius 비디오 보드는 비디오 입력을 가능하도록 비디오 리소스를 초기화시키고 이를 이용하여 비디오를 입력받아 연속적으로 텍스처 메모리에 업로드 시킴으로써 렌더링 모듈에서 입력받은 비디오를 가상환경의 렌더링 과정에 반영하도록 하여준다. 그림 7에서는 가상환경에서 참여자들이 서로 만나서 단어 맞추기 게임을 수행하는 장면에서 알파 값의 적용에 따른 영상의 변화를 적용 전과 후의 경우에 대하여 보여주고 있다. 블루 룸의 참여자가 책상 앞에 앉아 있도록 하기 위해서 물리적인

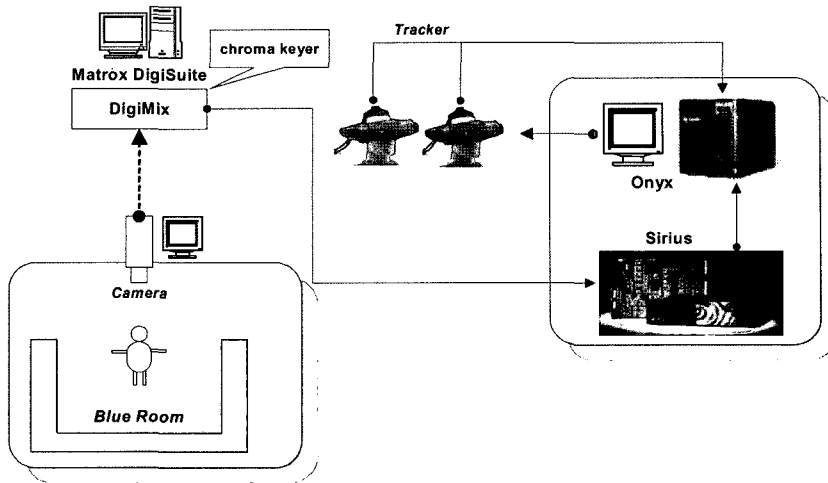
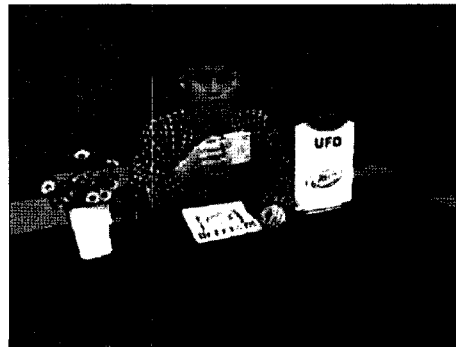


그림 6 시스템의 구성도



(a) 알파 값의 적용 전의 상태



(b) 알파 값의 적용 결과

그림 7 알파 값에 따른 비디오 텍스처의 변화

책상을 배경 색과 동일한 색으로 가리게 된다. 이러한 책상이나 상자 등의 물리적 소품을 이용하여 가상환경과 실제 환경의 상호작용을 유사한 형태로 연출하게 된다. 참여자는 이러한 소품들을 이용하여 가상환경에서 손으로 책상을 접촉하는 효과를 얻을 수 있다. 렌더링된 가상환경을 보며 상호작용을 하기 위해 블루 룸의 참여자에게 가상환경을 제시하는 별도의 디스플레이 장치를 사용한다.

본 논문에서 제안된 방법을 이용하여 비디오 아바타의 파라미터를 구하는 과정을 그림 8에서 보여주고 있다. 먼저, 그림 8(a)에서는 실제 카메라로 촬영된 앵커A의 목표 영상이다. 그림 8(b)는 크로마키더 가상공간 안에 존재하는 초기에 추정된 비디오 아바타의 위치를 보여준다. 비디오 아바타의 초기 위치는 경우에 따라 다르며 본 실험의 경우에는 참여자가 위치할 책상 중앙의 끝 부분에 위치하도록 하였다. 그림 8(c)는 그림 8(a)와

그림 8(b)의 영상을 50%씩 겹치게 함으로써 초기 위치의 차이를 보여주고 있다. 그림 8(d)는 제안된 방법으로 구해진 비디오 아바타의 최적의 위치를 보여준다. 그림 8(e)는 그림 8(a)와 그림 8(d)를 각각 50%씩 겹치게 표시함으로써, 구해진 비디오 아바타 파라미터의 정확도를 가늠해 볼 수 있다. 그림 8(f)는 구해진 비디오 아바타의 파라미터를 이용하여 가상 환경에 삽입된 비디오 아바타를 보여준다. 그림 9에서는 각 반복 수행에 따라 에러가 감소하는 추세를 그래프로 나타내었다. 이 그래프에서 x축은 반복수행 횟수를 나타내며 y축은 에러, 즉 목표영상과 현재 합성된 영상의 차이를 나타낸다.

표 1에서는 비디오 아바타 A에 대하여 파라미터 값의 초기위치와 구해진 최적위치를 비교하고 있다.

실험결과에서 알 수 있듯이 비디오 아바타는 가상환경에서 정확히 정합되어 마치 가상환경에 존재하는 듯한 느낌을 생성하게 된다. 이러한 결과는 제안된 방법으

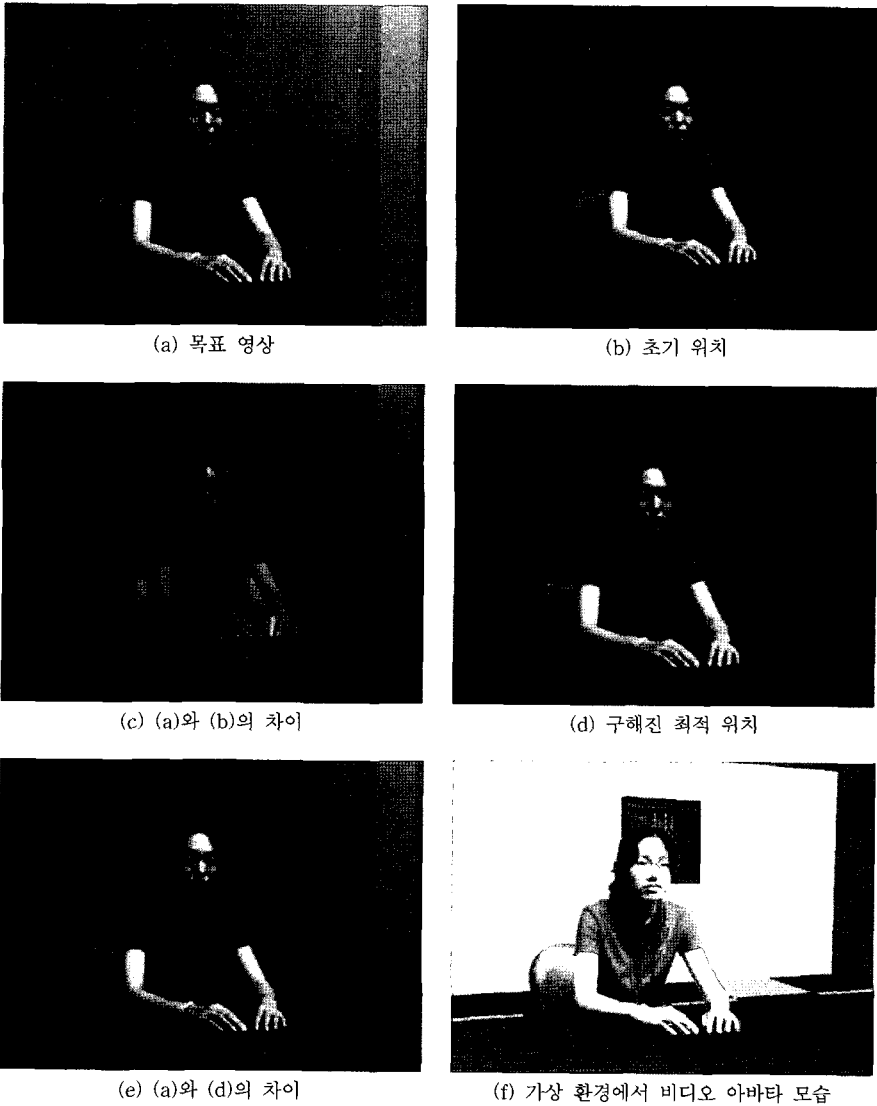


그림 8 비디오 아바타 A의 조정과정

로 구해진 정확한 비디오 아바타 파라미터를 적용했기 때문에 가능하다. 만약 비디오 아바타의 위치와 방향을 수작업으로 탐색하게 되면 비록 유사한 값을 찾을 수 있다 하더라도 정확한 값을 추출하는 것은 사실상 불가능하다. 실제 협업을 위한 렌더링 영상은 고해상도의 영상으로 렌더링 하지만, 준비 과정에서 비디오 아바타 파라미터를 구하는 과정에서는 320×240 크기의 영상을 이

용하여 시간을 단축하였다. 비디오 아바타 파라미터를 구하기 위해서는 5분정도의 시간이 소요되며 종료조건을 만족하지 않더라도 여러 그래프에서 급속한 에러 감소 부분을 지나는 지점에서 종료를 하여도 렌더링의 차이는 사람의 눈으로 크게 느껴지지 않기 때문에 수행시간을 적정히 조정하는 것이 가능하다.

일반적으로 충분한 시간적 여유를 가지고 정합과 리

표 1 비디오 아바타의 초기위치와 최적위치 비교

종류	구분	파라미터 값: x, y, z, h(heading), p(pitch), r(roll)
비디오 아바타A	초기위치	x=-0.17, y=94.75, z=45.12, h=0.01, p=0.99, r=-0.31
	최적위치	x=-10.22, y=92.64, z=32.79, h=0.23, p=-0.24, r=0.16

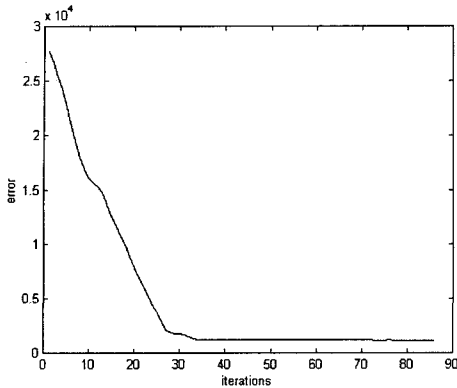


그림 9 비디오 아바타 A의 에러 그래프

허설을 수행한 후 협업을 수행하였다. 비디오 아바타의 파라미터를 구하는 과정은 리허설 과정에서 한번 수행되며 작업도중에 참여자의 위치와 방향이 다소 변경되더라도 책상 주변에 앉아서 작업하는 경우에는 큰 어려움 없이 작업을 진행할 수 있었다.

## 5. 결론 및 향후 연구 과제

가상현실 기술의 발전으로 컴퓨터가 만들어낸 가상의 환경에 대한 사실적 표현을 통하여 인간에게 다양한 가상 환경을 설득력 있게 제시하는 것이 가능하게 되었다. 본 논문에서는 현실 세계와 가상 세계의 요소들을 혼합하여 컴퓨터 그래픽으로 구현하기 어려운 사람의 몸동작과 얼굴 표정 등을 사실적으로 표현 가능한 비디오 아바타를 이용하였다. 배경이 전경과 분리되어 투명하게 입력되는 실시간 비디오 아바타를 가상 환경에 정합하기 위하여 실제 카메라와 가상 카메라의 특성을 같도록 조정하였고 실제 카메라와 가상 카메라가 바라보는 내용이 가장 유사하게 될 때 적용된 비디오 아바타의 파라미터를 최적의 값으로 결정하였다. 실험을 통하여 제안된 방법이 수작업으로 정합하던 방법의 한계를 극복하고 정확한 비디오 아바타 파라미터를 추출할 수 있음을 확인하였다.

향후에 진행될 주요 연구 과제로는 비디오 아바타 표현 방식에서 2차원 평면의 한계를 극복하기 위하여 여러 개의 카메라를 사용하여 보는 위치에 따라 카메라를 변환하는 방식의 도입이나 2.5차원 이상의 비디오 아바타 구현 방식을 적용하여 구현된 시스템에서 좀더 사실적으로 참여자를 표현하는 것이다. 또한 책상 주변 이외에 다양한 환경으로 확장에 대한 연구와 참여자간의 시선 맞춤, 응시 인식, 공간적 암시 등의 상호작용 요소들을 풍부하게 제공하기 위한 부가적인 방법을 추가하는 것이다.

## 참고 문헌

- [1] Chris Greenhalgh and Steve Benford, "Supporting rich and dynamic communication in large scale collaborative virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 1, pp. 14-35, 1999.
- [2] Jongwook Jin, Moonho Park, Heedong Ko and Hyeran Byun, "Immersive Telemeeting with Virtual Studio and CAVE," *Proceedings of International Workshop on Advanced Image Technology*, pp. 15-20, Daejeon, Korea, February 2001.
- [3] Ronald Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments* vol. 6, no. 4, pp. 355-385, 1997.
- [4] Paul Milgram and Fumio Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, Special issue on *Networked Reality*, vol. E77-D, no. 12, pp. 1321-1329, 1994.
- [5] Simon Gibbs, Costas Arapis, Christian Breiteneder, Vali Laloti, Sina Mostafawy and Josef Speier, "Virtual Studios: An Overview," *IEEE Multimedia*, vol. 5, no. 1, pp. 18-35, 1998.
- [6] Neal Stephenson, "Snow Crash," Bantam Books, 1992.
- [7] Tetsuro Ogi, Toshio Yamada, Yuji Kurita, Yoichi Hattori and Michitaka Hirose, "Usage of Video Avatar Technology for Immersive Communication," *Proceedings of First International Workshop on Language Understanding and Agents for Real World Interaction*, pp.24-31, Sapporo Japan, July 2003.
- [8] Youngjung Suh, Dongpyo Hong and Woontack Woo, "2.5D Video Avatar Augmentation for VRPhoto," *Proceedings of International Conference on Artificial Reality and Telexistence*, vol.12, pp.182-18, Tokyo Japan, Dec 2002.
- [9] Raphael Grasset, Julian Loose and Mark Billinghurst, "OSGARToolkit: Tangible+Transitional 3D Collaborative Mixed Reality Framework," *Proceedings of International Conference on Artificial Reality and Telexistence*, vol. 15, Christchurch New Zealand, Dec 2005.
- [10] Namgyu Kim, Woontack Woo and Makoto Tadenuma, "Photo-realistic Interactive 3D Virtual Environment Generation Using Multiview Video," *Proceedings of SPIE PW-EI-VCIP'01*, vol. 4310, pp. 245-254, Jan 2001.
- [11] Max Rathaler, "Virtual Studio Technology: An overview of the possible applications in television programme production," *European Broadcasting Union Technical Review*, no 268, Summer 1996.
- [12] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery, "Numerical recipes in C: The art of scientific computing,"



Cambridge University Press, 1992.

- [13] Roger Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE Journal of Robotics and Automation, vol. RA-3, no.4, pp. 323-344, 1987.
- [14] James D. Foley, Andries van Dam, Steven K. Feiner and John F. Hughes, "Computer Graphics: Principle and Practice," Addison-Wesley Publishing Company, 1990.
- [15] Mason Woo, Jackie Neider and Tom Davis, "OpenGL Programming Guide, Second Edition: The Official Guide to Learning OpenGL," Version 1.1, Addison-Wesley, 1997.
- [16] George Eckel, Ken Jones and Tammy Domeier, "OpenGL Performer Getting Started Guide," Silicon Graphics Incorporated Technical Publications Series, November 2000.



박 문 호

2002년 연세대학교 대학원 컴퓨터과학과 박사. 2002년~2003년 아이와대학교 가상현실응용센터 post-doc. 2003년~현재 삼성전자 연구원. 관심분야는 가상현실, 컴퓨터그래픽스, 컴퓨터비전



고 회 동

1989년 일리노이대학교 전산학과(박사)  
1988년~1990년 조지메이슨대학교 전산학과 객원 조교수. 1990년~현재 한국과학기술연구원 책임연구원. 2001년~현재 한국과학기술연구원 영상미디어연구센터장. 관심분야는 가상현실, 인공지능, HCI



변 혜 관

1980년 연세대학교 수학과 졸업(이학사)  
1983년 연세대학교 대학원 수학과 졸업(이학석사). 1987년 University of Illinois at Chicago, Computer Science(M.S.). 1993년 Purdue University, Computer Science(Ph.D.). 1994년~1995년 한림대학교 정보공학과 조교수. 1995년~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 인공지능, 영상인식, 영상처리