

## 다중사용자 위치추적을 위한 실루엣 기반 헤드 디텍션<sup>+</sup>

박지영<sup>0,\*</sup>, 이선민\* 김명희<sup>\*,\*\*</sup>

\*이화여자대학교 컴퓨터학과

\*\*이화여자대학교 컴퓨터 그래픽스/가상현실 연구센터

{lemie<sup>0</sup>, blue}@ewhain.net, mhkim@ewha.ac.kr

### Silhouette-based Head Detection for Tracking Multiple Users

Jiyoung Park<sup>0,\*</sup>, Seon-Min Rhee\* Myoung-Hee Kim<sup>\*,\*\*</sup>

\*Dept. of Computer Science & Engineering, Ewha Womans University

\*\*Center for Computer Graphics & Virtual Reality (CCGVR), Ewha Womans University

#### 요약

본 연구에서는 영상획득 조건을 충분히 만족시키지 못하는 몰입형 디스플레이 환경에서 다수의 사용자 머리 위치를 탐지하는 방법을 제안한다. 본 기법은 몰입형 가상환경에서 적외선 반사영상을 획득하고 그로부터 배경을 제거함으로써 얻어진 전경(foreground) 영역으로부터 프로젝션 히스토그램을 생성하고 사용자 실루엣을 추출하게 된다. 모든 사용자의 머리는 다각형 근사된 실루엣과 프로젝션 히스토그램에 기반하여 탐지된다. 또한 향후 몰입형 가상환경에서의 다중 사용자 트래킹을 지원하기 위해 스테레오 영상에서 탐지된 머리를 기준으로 탐색영역을 정의, 대응점을 결정하고 그에 기반하여 각 사용자 머리의 3차원 위치를 계산하였다.

#### 1. 서론

가상환경에서 사용자의 위치는 가상객체와 사용자간의 인터랙션을 설계하고 구현하는데 필수적인 정보이다. 실제로 사용자를 탐지하고 추적하기 위한 많은 시도들이 이루어지고 있는데, 특히 부가적인 마커나 장비의 착용이 불필요한 비전기반 사용자 탐지는 가상세계에서 이질감 발생을 최소화하고 사용자에게 편안하고 직관적인 인터랙션 방법을 제공할 수 있다는 장점으로 인해 그 연구가 활발하다. 또한 디스플레이 하드웨어의 발전으로 대형 스크린에 기반한 몰입형 디스플레이 장비가 가상현실 분야에서 널리 사용되고 있다. 이처럼 가상세계의 규모가 커지고 사용자의 수도 증가함으로써 대형 가상환경에서 다수의 사용자 위치를 탐지해내는 기술은 현재 구축된 몰입형 가상환경의 사용성을 극대화할 수 있다는 점에서 상당히 중요한 이슈이다. 그럼에도 불구하고 몰입형 가상환경에서 다중사용자 탐지에 관해 진행되는 연구는 거의 없으며 비전에 기반한 연구 또한 찾아보기 힘든 실정이다. 최근 Vorozcovs[1]는 6면의 완전 몰입형 디스플레이 내의 사용자의 위치를 알아내기 위한 광학 기반 기법을 제안하였으나 이 기법은 단일 사용자를 대상으로 하며 사용자의 머리부위에 빛을 반출하는 특수 장비를 장착하는 것을 요구한다. 이밖에 평면의 대형 스크린 환경에서 사용자 인터랙션을 지원하기 위한 얼굴인식 기법들이 다수 연구된 바가 있다. 그러나 몰입형 가상환경에서는 사용자의 얼굴

이 잘 나타나는 영상을 얻기가 상당히 어렵다. 먼저, 사방을 둘러싼 스크린 때문에 카메라 설치 위치가 극히 제한적이다. 또한 카메라가 설치되더라도 디스플레이를 위해 주변조명이 차단되고 프로젝션되는 빛만이 허용되기 때문에 영상획득을 어렵게 하는 불안정한 조영조건이 형성된다. 실제로 4면 구성 CAVE<sup>TM</sup>-like 시스템에서는 스크린 상단에 카메라를 설치하는 것이 가능하나 그 위치에서 획득된 카메라 영상에서는 얼굴에서의 특징부위가 잘 나타나지 않는 경우가 대부분이다. 비전기반 얼굴인식 기법에서는 사용자 얼굴의 특징부위를 추출하는 것이 기본이 되기 때문에 이를 몰입형 가상환경에 그대로 적용해서는 안정적인 사용자 위치 탐지 결과를 얻을 수 없다.

이에 본 연구에서는 영상획득 조건을 충분히 만족시키지 못하는 몰입형 디스플레이 환경에서 다수의 사용자 머리 위치를 탐지하는 방법을 제안한다. 본 기법은 몰입형 가상환경에서 적외선 반사영상을 획득하고 그로부터 배경을 제거함으로써 얻어진 전경(foreground) 영역으로부터 프로젝션 히스토그램을 생성하고 사용자 실루엣을 추출하게 된다. 사용자 실루엣은 다각형 근사를 통해 단순화되고 그것과 프로젝션 히스토그램에 기반하여 다중사용자의 머리가 탐지된다. 또한 본 연구에서는 향후 몰입형 가상환경에서의 다중사용자 트래킹을 지원하기 위해 스테레오 영상을 획득하였고 추출된 머리위치를 기준으로 탐색영역을 정의, 대응점을 결정하고 그에 기반하여 각 사용자 머리의 3차원 위치를 계산하였다.

<sup>+</sup> 본 연구는 정보통신부 대학정보통신연구센터(ITRC) 육성지원사업 지원에 의해 수행되었음.

2. 적외선 반사 영상에서의 사용자 실루엣 추출

조명조건이 불안정한 프로젝션 기반 몰입형 가상환경에서 효과적으로 사용자 실루엣을 얻어내기 위해 적외선 반사 영상을 획득한다. 본 연구에서는 실험 환경으로 정면, 좌/우측면, 바닥면의 총 4면 스크린으로 구성된 CAVE™-like 시스템을 사용하였다.

2.1 적외선 반사 영상 생성

CAVE™-like 시스템에 설치된 카메라와 적외선 광원의 위치는 그림 1과 같다. 본 연구에서는 전면 스크린 상단에 적외선 광원 2개와 적외선 필터를 장착한 일반 흑백카메라를 설치하였다. 사용자의 전신에 적외선이 고루 분산될 수 있도록 2개의 광원이 각각 사용자의 상반신과 하반신을 향하도록 하였다. 또한 보다 정확한 배경제거 결과를 위하여 케이브 입구에 적외선 흡수재질의 커튼을 설치하였다. 적외선 투과 필터를 통해 얻어지는 흑백영상에서는 사용자 영역이 다른 영역보다 상대적으로 밝게 나타나게 되며 가시광선은 필터에 의해 걸러지므로 카메라에 입력되지 않는다. 따라서 적외선 반사 영상은 정적배경을 전제로 하는 배경제거 알고리즘에 적용할 수 있다. 본 연구에 사용된 적외선 광원은 Computar사의 IR75 Infrared Lamp (750nm, 2개)이며, 적외선 투과 필터는 Edmund Optics 사의 사각형태의 FILTER OPTCAST IR이다. 흑백 카메라는 Point Grey Research사의 Dragonfly를 사용하였다.

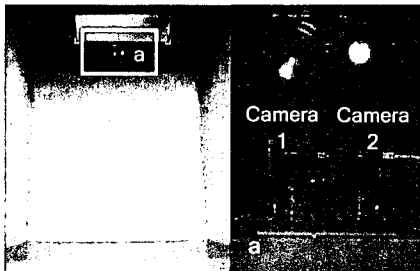


그림 1 적외선 반사영상 획득을 위한 하드웨어 셋업

2.2 배경제거를 통한 사용자 실루엣 추출

적외선 반사 영상에서 사용자를 추출하기 위해 [2]에서 제시된 배경제거 기법을 적용하였다. 적용된 기법은 전처리 단계인 배경흡득 단계, 배경분리 단계, 후처리 단계로 구성되며 배경흡득 단계에서는 사용자가 존재하지 않는 배경 영상의 n개 연속 프레임에 대해 각 픽셀 위치의 n개 값에 대한 평균, 표준편차가 계산된다. 배경분리 단계에서는 사용자를 포함한 연속 영상이 입력되며 각 픽셀값과 배경흡득 단계에서 계산된 동일 위치상의 통계값을 비교함으로써 변화 여부를 판별한다. 후처리 단계에서는 팽창(dilation)과 미디언 필터를 적용하여 사용자 실루엣을 보다 자연스럽게 나타나게 한다.

3. 실루엣 기반 다중 사용자 머리 탐지

다수의 사용자의 머리를 탐지하기 위한 본 연구의 기법은 Hydra[3]에서 착안한 것으로 사용자 실루엣과 프로젝션 히스토그램을 사용하였다. 현재로서는 사용자의 수를 사전에 입력하여 수행시간을 줄였고 3차원 머리 위치 계산을 위한 대응점은 cross correlation에 기반하여 결정하였다.

3.1 다중 사용자 머리 탐지

실루엣 영상으로부터 프로젝션 히스토그램을 생성하고 실루엣은 다각형 근사를 통해 단순화시킨다. 프로젝션 히스토그램은 배경제거의 결과인 이진영상에서 가로축의 모든 위치에 대해 세로축의 모든 픽셀을 검사하여 그 값이 0이 아닌 픽셀의 총 수를 나타내게 된다. 즉, 이진영상에서 특정 가로축 위치 x를 따라 세로축상에 0이 아닌 픽셀이 많을 경우, x위치에서의 히스토그램 값이 크게 나타난다. 실루엣의 다각형 근사는 Intel OpenCV 라이브러리 [6]에서 제공하는 모듈을 사용하였으며 이는 단순한 파라미터 입력을 통해 근사 정도를 달리해서 단순화시킨 실루엣을 얻을 수 있다. 다음으로 단순화된 실루엣에 코너 탐지를 적용하게 되며 이 때 위쪽으로 돌출된 모서리 부위가 머리부위로 결정된다. 탐지된 머리의 수가 입력된 사용자수보다 많은 경우 프로젝션 히스토그램의 값을 참조하여 최종 머리 위치를 결정한다. 일반적으로 머리는 사용자 영역의 상위에 위치함과 동시에 프로젝션 히스토그램에서는 평균값보다 큰 값을 가진다. 다시 말해, 가로축에서 특정 위치가 사용자의 머리 즉, 몸통 근처인 경우의 히스토그램 값은 팔 위치에서의 히스토그램 값보다 크다. 이와 같은 히스토그램의 참조를 통해 사용자가 손을 옆으로 들었을 경우 손이 머리로 탐지되는 오류를 피할 수 있다. 다각형 근사된 사용자 실루엣과 프로젝션 히스토그램의 예는 그림 2와 같다.



그림 2 사용자 실루엣의 다각형 근사(좌) 및 프로젝션 히스토그램(우)의 예

3.2 스테레오 기반 3차원 머리 위치 계산

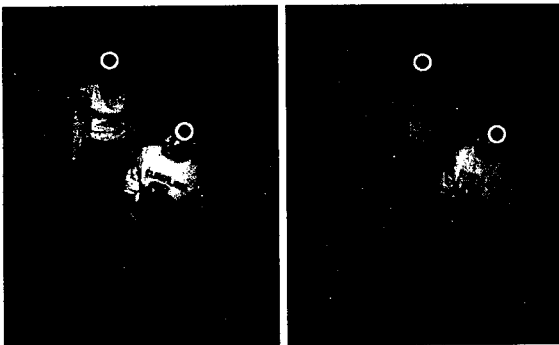
다음으로 각 사용자 머리의 정확한 3차원 위치를 계산한다. 3차원 위치 계산은 향후 개별 사용자의 위치를 추적하여 인터랙션을 디자인하는데 필수적인 작업이다. 현재까지 머리탐지 단계를 통해 스테레오 영상에서 각 사용자의 머리끝에 대한 두 개의 정점을 결정하였다. 그러나 이 위치는 단순화된 실루엣에 기반한 것이므로 이를 사용

해서는 정확한 3차원 머리 위치를 계산할 수 없다. 본 연구에서는 스테레오 영상에서 탐지된 머리 위치를 기준으로 대응점을 탐색하기 위한 탐색 영역을 정의하고 cross correlation에 기반하여 대응점을 결정하였다. 즉, 왼쪽 영상의 탐색영역에 대해 마스크를 설정하고 오른쪽 영상에서 같은 크기의 마스크를 적용하여 두 개의 마스크로부터 cross correlation을 계산한다. 양쪽의 마스크의 위치를 조금씩 이동시키면서 correlation 값을 계산하고, 이것이 가장 큰 경우의 양쪽 마스크의 중심점이 최종 대응점으로 결정된다.

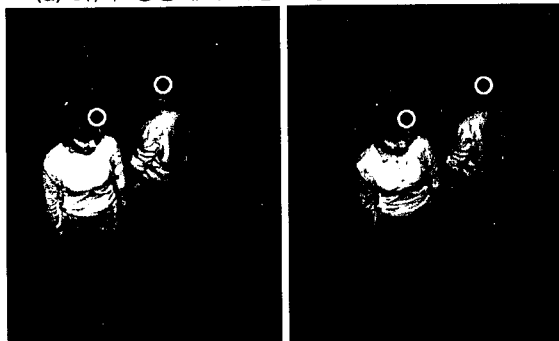
4. 실험결과

CAVE™-like 시스템에서 2명의 사용자 머리 탐지를 실험한 결과는 그림 3과 같다. (a)와 (b)는 서로 다른 시점에서 획득된 좌/우 카메라 영상에서 추출한 사용자 영역을 나타낸다. 모든 영상에서 모든 사용자의 머리부위가 정확히 탐지되었으며 그림에서 표시된 흰색 원의 중심이 3차원 머리위치 계산을 위한 대응점이 된다.

실험은 dual 2.8GHz Xeon 프로세서 PC에서 수행되었다. 수행속도는 몰입형 디스플레이 환경에 존재하는 사용자의 수, 탐색영역과 마스크의 사이즈에 따라 달라지는데, 2명의 사용자일 경우 탐색영역과 마스크의 사이즈를 각각 40X40, 25X25 픽셀로 설정하였을 때 약 20Hz~25Hz로 나타났다.



(a) 좌/우 영상에서 다중 사용자 머리 탐지 결과 1



(b) 좌/우 영상에서 다중 사용자 머리 탐지 결과 2  
그림 3 두 시점 (a), (b)에서의 다중 사용자 머리 탐지

5. 결론 및 향후 연구

본 논문에서는 몰입형 가상환경 내에 존재하는 다수 사용자의 머리 위치를 탐지하는 기법을 제시하였다. 몰입형 가상환경에서 여러 명의 사용자를 지원하는 탐지나 트래킹, 인터랙션 기법의 연구가 활발하지 않은 가운데 본 기법은 현재 구축된 가상환경의 사용성을 극대화할 수 있는 가능성을 제시한다. 또한 제안된 기법은 마커와 같은 추가적인 도구가 불필요하며 정확하고 안정적인 탐지 결과를 제공함으로써 가상세계에서의 사용자 몰입감을 최대화하고 자연스러운 인터랙션의 구현을 가능하게 한다. 또한 스테레오에 기반하여 사용자의 3차원 위치를 계산할 수 있도록 하였으며 빠른 수행속도를 제공한다.

현재 본 기법은 입력영상에서 사용자들이 서로의 많은 부분을 가리는 경우 만족스러운 탐지 결과를 제공하지 못한다. 그러므로 향후 연구로서 이를 개선하기 위한 추가적인 알고리즘을 개발할 필요가 있다. 또한 제안 기법에 기반하여 가상환경에서 여러 명의 사용자 각각을 인식하고 구별할 수 있는 다중 사용자 트래킹 기법을 제시하고자 한다. 마지막으로 몰입형 가상환경에서의 추가적인 장비 장착 없이 가능한, 다중 사용자를 위한 인터랙션을 디자인하고 그를 위한 어플리케이션을 개발할 예정이다.

참고문헌

- [1] A. Vorozcovs, A. Hogue, W. Stuerzlinger, " The Hedgehog: A Novel Optical Tracking Method for Spatially Immersive Displays", In IEEE VR 2005, 2005.
- [2] 박지영, 이선민, 김명희, " 혼합현실환경을 위한 능동 적외선 기반 동적배경 제거", 2004 한국정보과학회 추계학술발표대회 논문집, 2004.
- [3] I. Haritaoglu, D. Harwood, and L. S. Davis, " Hydra: Multiple people detection and tracking using silhouettes", In IEEE International Workshop on Visual Surveillance, 1999.
- [4] I. Haritaoglu, D. Harwood, and L. Davis., " W4: Who, when, where, what: A real time system for detecting and tracking people", In Third Face and Gesture Recognition Conference, 1998.
- [5] I. Haritaoglu, D. Harwood, and L. Davis., " Ghost: A Human Body Part Labeling System Using Silhouettes", In International Conference on Pattern Recognition, 1998.
- [6] Intel OpenCV library,  
<http://www.intel.com/research/mrl/research/opencv>