

# 적외선 카메라를 이용한 실시간 사용자 분리 시스템\*

박영민, 우운택  
광주과학기술원 U-VR 연구실  
{ypark,wwoo}@gist.ac.kr

## Real-time user subtraction system using an IR camera

Youngmin Park, and Woontack Woo  
GIST U-VR Lab.

### 요약

본 논문에서는 적외선 카메라를 이용한 사용자 분리 시스템을 제안한다. 사용자를 배경으로부터 실시간에 분리하는 것은 컴퓨터와 사용자간의 시각 기반 상호작용에 필수적이다. 블루스크린으로 대표되는 컬러 기반 분리 시스템은 고가의 장비를 필요로 하고, 확률 기반의 기법은 카메라가 고정된 경우에만 가능하며, 공통적으로 두 방법 모두 조명 등 주변 환경의 영향에 민감하다. 이러한 문제점을 극복하기 위해 본 논문에서는 적외선 카메라 기반의 시스템을 제안한다. 제안된 시스템은 적외선 카메라와 일반 카메라, 적외선 반사체 스크린으로 구성된다. 적외선 카메라는 스크린으로부터 영상 내 사용자의 마스크를 생성하는데 사용된다. 적외선 카메라에 고정되어있는 일반 카메라는 두 카메라간의 시야 영역의 오차를 보정하여 마스크 영역에 해당하는 컬러 텍스처를 획득한다. 따라서, 본 시스템은 주변 환경의 영향에 민감하지 않고 강건한 사용자 분리가 가능하다. 제안된 시스템은 비전 기반 제스처 인식, 혼합현실 등의 응용 분야에 활용될 수 있다.

Keyword : Subtraction, Vision-based Interaction, HCI, IR camera

## 1. 서론

컴퓨터의 연산속도와 통신기술, 렌더링 기술이 급속도로 발전함에 따라 사용자가 사용할 수 있는 콘텐츠가 다양화 되고, 컴퓨터에 의해 생성된 가상공간과 관련된 다양한 분야에 대한 연구가 활발히 진행되고 있다. 이러한 다양한 콘텐츠를 대규모의 디스플레이 장치를 통해 사용자에게 제공할 경우 기존의 키보드와 마우스를 사용한 상호작용 방식은 사용자의 몰입감을 떨어뜨리고, 사용자의 의도전달이 힘들다 [1][2].

대규모 콘텐츠나 사용자와의 상호작용을 효율적으로 지원하기 위해 사용자의 제스처 인식과 관련된 다양한 연구가 이루어지고 있다. 영상 기반 인터페이스의 경우 전체 영상에서 관심 영역만을 추출하는 배경 추출 기법은 필수적이다.

배경 추출 기법은 크게 블루스크린으로 대표되는 크로마 키 (Chroma Key) 기반의 추출기법과, 일정시간 동안의 학습과정을 필요로 하는 확률 기반의 기법으로 나눌 수 있다. 블루스크린 시스템은 움직이는 영상으로부터 실시간으로 배경 추출이 가능하고 추출 결과의 품질이 좋은 반면에 시스템을 구축하는데 고가의 장비가 필요하고, 색 번짐 효과가 발생하며,

\* 본 연구는 한국정보통신대학교 디지털미디어연구소의 프로젝트와 광주과학기술원 실감콘텐츠연구센터(ICRC)에 의하여 지원되었음

스크린과 동일한 색상의 사용이 불가능하다 [3]. 확률 기반의 기법은 조명, 물체나 카메라의 이동 등의 환경 변화에 취약하다 [4][5].

본 논문에서는 이러한 문제점들을 극복하고 실시간으로 배경에서 사용자를 추출하기 위한 시스템을 제안한다. 제안된 시스템은 적외선 카메라, 일반 카메라, 그리고 적외선 반사체로 제작된 스크린으로 구성된다. 제안된 시스템의 동작 과정은 다음과 같다. 우선, 적외선 카메라와 일반 카메라를 정밀하게 캘리브레이션한다. 그런 다음, 적외선 카메라를 이용하여 스크린과 카메라 사이에 위치한 배경과 분리된 사용자 마스크를 생성한다. 마지막으로, 생성된 마스크의 텍스처 정보를 얻는다.

본 논문에서 제안된 시스템은 실내 환경에서의 조명의 영향에 민감하지 않고 실시간 사용자 분리가 가능하며, 정확한 사용자 분리 결과를 제공한다. 주변 환경의 영향에 민감하지 않으므로 스크린의 크기에 따라 이동이 가능하도록 제작될 수 있는 장점이 있다. 제안된 시스템은 비전 기반 제스처 인식, 혼합현실 등의 응용 분야에 활용될 수 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 제안된 시스템의 구성을 설명하고, 3 장에서는 실험결과를 기술한다. 4 장에서는 결론과 향후 연구방향에 대해 언급한다.

## 2. 제안된 시스템 구성

제안된 시스템은 다음과 같이 세 단계로 구성된다. 첫 번째 단계는, 적외선 카메라와 일반 RGB 카메라의 캘리브레이션이다. 적외선 카메라와 RGB 카메라에서 획득한 영상의 특정 영역을 일치시키기 위해서는 두 카메라간의 거리를 측정해야 한다. 두 번째 단계는 적외선 카메라를 이용한 반사체 스크린 내 사용자의 마스크 획득이다. 적외선 카메라에서 획득한 영상은 사용자의 마스크를 추출하기에 용이하다. 그러나 반사체 스크린의 크기를 고려하여 반사체에 해당하는 영역이 영상 전체를 차지하지

않는 경우 반사체 이외의 모든 영역이 아닌 특정 부위만을 마스크로 결정해야 한다. 세 번째 단계는 일반 RGB 카메라를 이용하여 사용자 마스크에 해당하는 영역의 텍스처 획득하는 단계이다. 적외선 카메라로부터 획득한 영상을 일반 카메라로부터 획득한 영상에 마스크로 적용한다. 그러나 두 카메라간의 거리에 의해 각 카메라의 가시범위에 오차가 생기므로 카메라와 사용자의 거리가 가까울수록 마스크와 실제 사용자의 형태가 일치하지 않는다. 따라서 이 단계에서는 미리 계산된 카메라의 캘리브레이션 정보를 이용하고 두 영상의 차이를 비교하여 마스크 영역에 해당하는 텍스처를 최대한 가깝게 획득한다. 그림 1 은 전체 시스템의 구성도를 보여준다.

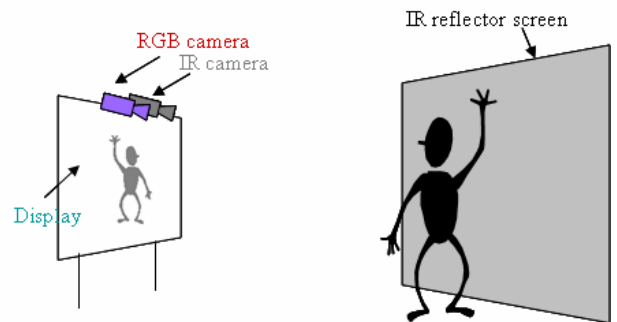


그림 1. 시스템 구성도

### 2-1. 적외선 카메라 기반 사용자 분리

본 논문에서 제안한 시스템은 위에서 설명한 특성을 이용하여 영상에서 사용자의 영역을 추출한다. 적외선 카메라는 특성상 블루스크린과 마찬가지로 적외선 반사체로 제작된 스크린을 필요로 한다. 블루스크린은 컬러 키 기반으로 컬러 영상에서 배경을 제거함으로써 사용자의 영역과 텍스처를 동시에 획득한다. 반면에 제안된 시스템은 적외선 카메라를 통해 획득한 흑백 영상에서 사용자의 영역에 해당하는 마스크만을 획득할 수 있다. 그러나 제안된 시스템은 관심 영역을 확률기반의 방법과는 다르게 학습과정 없이 명확히 구분할 수 있으며, 블루스크린과 달리 가시광선에 해당하는 조명의 영향을 받지 않는다. 마스크 영역에 해당하는 텍스처 정보를

추출하기 위해서는 적외선 카메라 이외에 컬러영상 획득을 위한 일반 RGB 카메라가 필요하다. RGB 카메라는 적외선 카메라에 고정되어 있지만, 두 카메라간의 거리가 떨어져 있고, 카메라의 종류에 따라 두 카메라의 가시영역이 정확히 일치하지 않는다. 그림 3 은 이것을 보인 것이다.

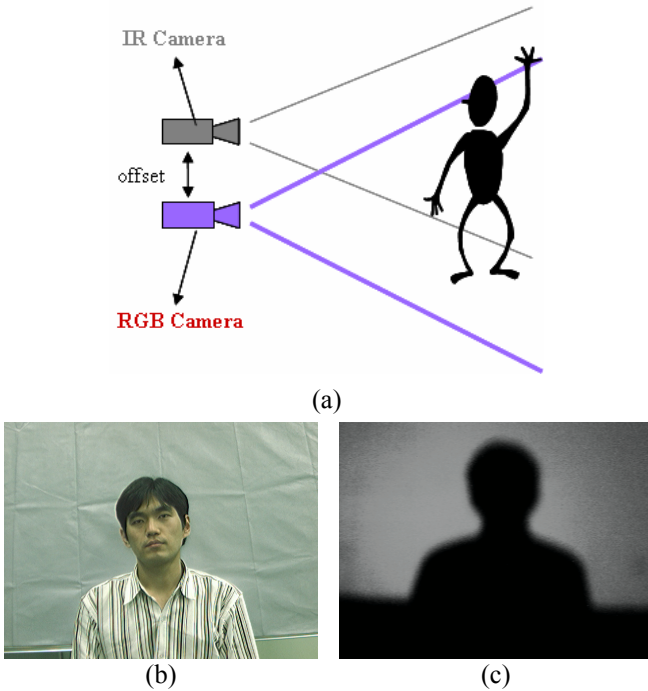


그림 3. 카메라간 가시영역의 오차 (a) 개념도 (b) 일반 RGB 카메라 영상 (c) 적외선 카메라의 영상

### 2-2. 시스템 컴포넌트

그림 4 는 시스템의 사용자 분리 처리과정을 컴포넌트 별로 보여준다. 각 컴포넌트의 세부적인 사항은 다음과 같다.

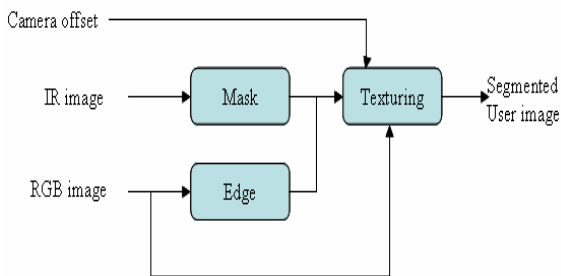


그림 4. 사용자분리 처리 과정

마스크 (Mask): 적외선 카메라로 획득한 영상으로부터 사용자에게 해당하는 영역의 마스크를 추출한다. 획득한 영상에 나타난 반사체 스크린의 상태에 따라 그림 5 와 같이 4 가지의 경우로 나눌 수 있다. 그림 5(a)는 반사체 스크린이 사용자 부분을 제외한 영상 전체에 나타난 경우이다. 그림 5 (b)는 반사체 스크린 하단 부 경계가 영상에 나타난 경우이다. 그림 5(c)는 반사체 스크린의 상단 부 경계가 영상에 나타난 경우이다. 그림 5 (d)는 반사체 스크린의 한쪽 모서리 부분만이 영상에 나타난 경우이다. 그림 5 의 (a), (c), (d)의 경우, 영상에 이진 임계값 (binary threshold)을 적용함으로써 마스크를 추출할 수 있다. (b)의 경우에는 단순히 임계값을 적용하는 것만으로는 사용자에게 해당하는 영역과 스크린 외부의 영역이 구분이 되지 않으므로, 스크린의 하단 부를 찾아내 마스크 추출을 위한 경계로 사용한다.

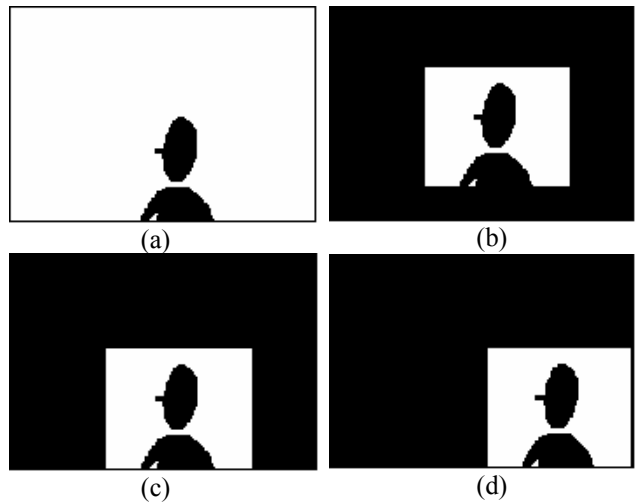


그림 5. 적외선 카메라의 영상: 영상에 나타난 반사체 스크린의 부위별 구분

테두리 (Edge): 일반카메라에서 획득한 영상으로부터 적외선 영상에서 추출된 마스크와 일치하는 영역을 예측하기 위해 테두리 영상을 추출한다. 테두리 영상을 획득하기 위한 방법으로 Canny edge 검출 알고리즘이 사용되었다 [6][7].

텍스처링 (Texturing): 여기에서는 마스크에

해당하는 텍스처를 컬러영상에서 추출한다. 두 카메라는 캘리브레이션되어 있지만, 카메라와 사용자의 거리에 따라 오차범위가 변하므로 부가적인 연산이 필요하다. 먼저 마스크의 경계 사각형 (bounding box)를 계산한다. 다음은 두 카메라간의 보정 정보를 이용하여 두 영상의 일치 영역을 예측한다. 예측 지점을 기준으로 마스크에서 계산한 것과 동일한 크기의 경계 사각형 (bounding box)을 테두리 영상과 비교하여 가장 이상적인 좌표를 찾는다. 경계 사각형의 네 개의 꼭지 점을 이용하여 두 영상 간 지정 영역의 호모그래피 (homography)를 계산한다. 두 영상간의 호모그래피는 다음과 같이 3×3의 행렬로 정의된다.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (1)$$

적외선 영상의 점 (pixel)  $(x, y)$  는 RGB 영상의 점  $(x', y')$  로 다음과 같이 변환된다 [8].

$$\begin{aligned} x' &= \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ y' &= \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{aligned} \quad (2)$$

마지막으로 적외선 영상의 각 픽셀 (pixel) 값을 컬러영상에 마스크로 적용하여 추출영역에 해당하는 텍스처 정보만을 획득한다.

### 3. 실험 결과

실험에는 DELL 650 (Intel Xeon 2.8 GHz dual CPU, 2Gb RAM)이 사용되었다. 적외선 카메라와 RGB 카메라로는 두 대의 UniBrain Fire-i 400 카메라가 사용되었다. 카메라의 해상도는 640×480 이고 초당 15 프레임을 얻는다.

그림 6 은 적외선 반사체로 제작된 스크린과 실험에 사용된 카메라가 설치되어 있는 것이다. 적외선 반사체 스크린은 육안으로 보기에는 정확한 평면이 아니고 완전히 평평하지 않지만,

표면의 재귀반사 (retro-reflective) 성질에 의해 적외선 카메라에는 전체가 거의 일정한 명암으로 나타난다.

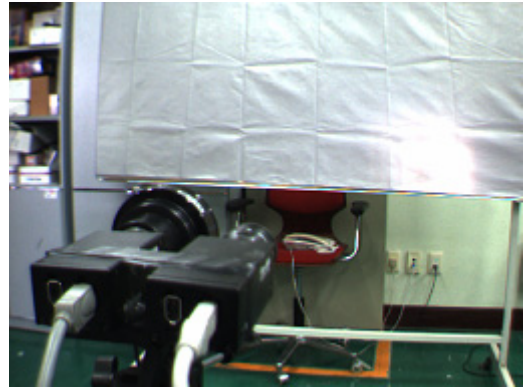


그림 6. 설치된 시스템 구성

그림 7 은 적외선 카메라와 RGB 카메라에서 획득한 영상과, 각 영상의 마스크, 경계 영상을 보여준다. 그림 7 (a)와 그림 7 (b)는 적외선 카메라와 RGB 카메라를 통해 획득한 영상을 보여준다. 그림 7(c)는 그림 7(a)에 이진 임계값을 적용하여 획득한 마스크를 보인 것이고, 그림 7 (d)는 (b)에서 추출된 테두리 영상이다.

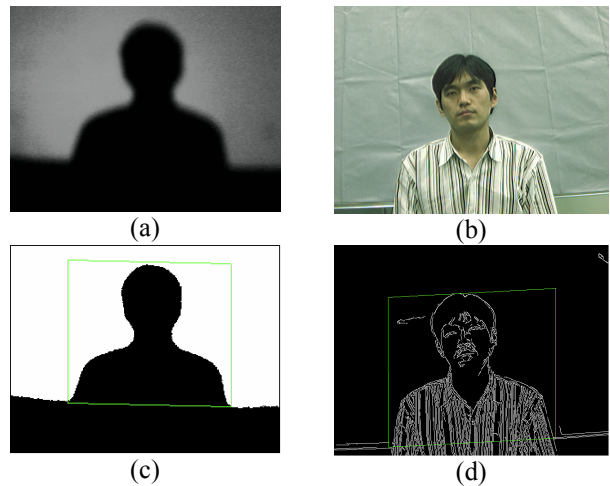


그림 7. (a) 적외선 카메라 영상 (b) RGB 카메라 영상 (c) 마스크와 경계 사각형 (d) 테두리 영상과 경계 사각형

그림 8 은 최종적으로 분리된 사용자의 영상을 보여준다. 마스크는 사용자의 형태와 대체적으로 일치하고, 노이즈가 없음을 알 수 있다. 그러나, 두 카메라 간의 거리로 인해 획득한

영상이 정확히 일치하는 영역을 나타내지 않고, 일치영역을 결정하는 과정에서 오류가 발생한다.



그림 8. 분리된 사용자 영상

#### 4. 결론 및 향후 연구

본 논문에서는 적외선 카메라를 이용한 사용자 분리 시스템을 제안하였다. 제안된 시스템은 적외선 카메라와 적외선 반사체로 스크린 앞에 위치한 사용자의 영역만을 강건하게 분리해냈다. 또한, 적외선 카메라와 고정되어있는 RGB 카메라로부터 분리된 영역에 해당하는 컬러정보를 추출함으로써, 조명 등의 주변 환경 변화에 영향을 받지 않았다. 따라서 제안된 시스템은 비전 기반 제스처 인식, 혼합현실 등의 응용 분야에 활용될 수 있다. 향후에는 두 카메라의 시야가 완전히 일치하지 않으므로 추출된 마스크와 텍스처가 정확히 일치하지 않는 문제를 보완하기 위해 마스크 영역과 텍스처 정보를 최적화하는 기법에 대한 연구를 진행할 예정이다.

#### 참고 문헌

[1] J. Segen and S. Kumar, "GestureVR: Vision-Based 3D Hand Interface for Spatial Interaction" , Proceedings of the sixth ACM international conference on Multimedia, pp.455-464, 1998.

[2] R. Sharma, V. I. Pavlovic and T. S. Huang, "Toward Multimodal Human-Computer Interface.", Proceedings of IEEE special issue on Multimedia Signal Processing, 86(5):pp. 853-869, 1998.

[3] Ben-Ezra, M. "Segmentation with invisible keying

signal", Proceedings of Computer Vision and Pattern Recognition, 2000. IEEE Conference on , Volume: 1 , pp.32 - 37 vol.1, 2000

[4] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection, ", Proceedings of IEEE ICCV'99 FRAME-RATE Workshop, Kerkyra, Greece, September 1999

[5] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric Model for Background Subtraction, " Proceedings of 6th European Conference on Computer Vision, Dublin, Ireland, June/July 2000

[6] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), pp. 679-698, 1986

[7] Intel OpenCV library., <http://www.intel.com/>

[8] R. Hartley and A. Zisserman, "*Multiple View Geometry in Computer Vision*", Cambridge University Press, March 2004.