

스테레오 카메라를 이용한 증강현실 시스템

*안재우 *고민수 *권혁민 *이동석 *유지상

광운대학교 전자공학과

*skai9@kw.ac.kr

Augmented Reality System using Stereo Camera

*An, Jae-woo *Ko, Min-soo *Kwon, Hyeok-min *Lee Dong-seok *Yoo, Jisang

Dept. Electronic Engineering Kwangwoon University

요약

본 논문에서는 스테레오 카메라를 이용하여 3차원 공간을 인식한 후 판별된 손동작에 의해 3차원 객체를 제어 할 수 있는 실시간 시스템을 구현 하고자 한다. 스테레오 카메라를 통해 획득한 현재 영상에서 배경 영상을 제거한 후 피부색의 특징을 이용한 이진화를 수행하여 손의 영역을 찾고 관심 영역에 대해서 변이(disparity)를 추정하여 손 영역의 깊이정보를 추가한 3차원 좌표를 실시간으로 얻을 수 있다. 손의 특징 점을 이용하여 손목을 찾고 손동작을 판별한다. 손을 관심 영역으로 설정한 시스템에서 3차원 공간 좌표는 실시간으로 사용자의 손의 움직임에 의해 획득되고 이를 가상과 현실이 합쳐진 3차원 증강현실과의 연동을 하게 한다. 실험을 통해 제안하는 실시간 시스템은 손과 카메라의 각도와 손가락의 개수에 따라 평균 손동작 인식률이 83%가 되는 것을 검증하였다.

1. 서론

오늘날 정보 산업분야에서 영상 및 음성 등에 관한 여러 가지 기반기술과 직접회로 기술, 신호 처리 기술, 다중화 기술 등이 급속하게 발전함에 따라 방대한 정보를 동시에 다룰 수 있게 되었고, 독립적으로 이루어지던 영상, 음성 데이터가 하나의 매체로 처리 및 전송되는 멀티미디어 시대가 도래함에 따라 사용자들은 끊임없이 보다 사실적이고 현장감 있는 영상을 요구하게 되었다. 우리가 살고 있는 세계가 3차원 공간임을 고려하여 입체감 및 현실감을 위한 3차원 입체 영상의 구현에 대한 새로운 연구 주제가 등장한 것이다.

이에 본 논문에서는 다시점 스테레오 카메라를 이용한 3차원 공간을 인식할 수 있는 시스템을 구현하고자 한다. 관심 영역을 분할하는 방법을 통해 손 영역 추출을 하는데 속도를 향상 시키는 알고리즘을 제안한다. 또한 관심 영역의 움직임을 3차원 좌표로 추적하여 사용자에게 제공한다. 깊이 정보를 포함한 3차원 물체를 실시간으로 사용자의 움직임에 따라 움직이므로써 직접 조작하는 것 같은 현실감을 준다. 실험을 통하여 추출된 손 영역과 동작 인식률을 측정된 결과 인식의 정확성을 입증하였다.

본 논문의 구성은 다음과 같다. 2장에서는 전체적인 구조도와 세부적인 동작을 설명하고, 3장에서는 제안한 손 영역 인식에 대한 측정하는 실험을 수행한다.

2. 시스템 구조도

그림 1에서는 시스템의 전체적인 흐름을 보여준다. 스테레오 카메라로부터 얻은 양안 시점의 영상에서 배경 영상을 제거한 후 피부색의 특징을 이용한 이진화를 수행하여 손 영역을 찾고 변이를 추정하여 손

영역의 깊이정보를 추가한 3차원 좌표를 찾는다. 손의 특징점을 이용하여 손목을 찾고 손동작을 판별한다. 이를 가상현실과 합쳐진 3차원 증강현실과의 연동을 하였다.

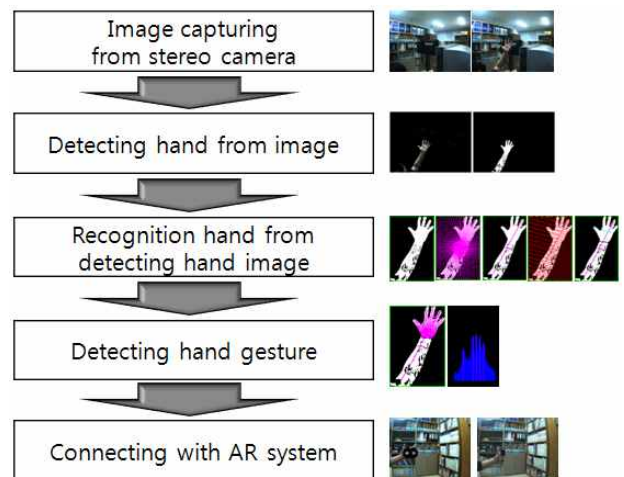


그림 1. 시스템 전체구성도

2-1. 관심 영역 분리

스테레오 카메라를 사용하여 획득된 최초 영상은 R, G, B, 투명도에 각 8bit씩 한 화소에 총 32bit를 가지는 1280×960 크기의 영상이다. 하지만 이를 실시간으로 처리하기에는 정보량이 너무 많기 때문에 가로, 세로 각각 4배씩 down-sampling을 하여 qVGA(320×240)급의 영상으로 변환하고 사용되지 않는 투명도 값을 제외하여 R, G, B 각

8bit씩 한 화소에 총 24bit를 갖도록 정보량을 줄여준다.

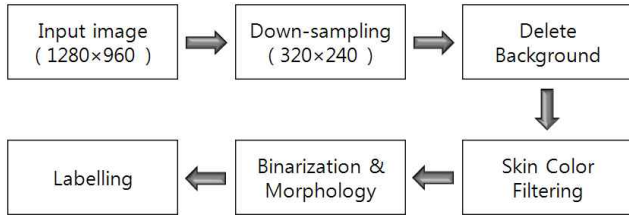


그림 2. 관심 영역 분리하는 흐름도

Down-sampling과 bit변환으로 얻는 효과는 식 (1)과 같이 약 21배 정도의 정보를 줄이는 효과를 얻을 수 있다.

그림 2. 관심 영역 분리하는 흐름도

$$\begin{aligned} \text{최초 영상: } 1280 \times 960 \times 4 &= 4915200 \text{ (Byte)} \\ \text{Down-sampling: } 320 \times 240 \times 3 &= 230400 \text{ (Byte)}. \end{aligned} \quad (1)$$

변환된 영상의 불필요한 정보 중 가장 많은 양을 차지하는 것은 배경 영상이다. 만약 영상 안에 물체가 들어오게 되면 물체가 배경을 가리게 된다. 따라서 물체에 의해 배경이 가려진 부분의 화소 값은 배경의 화소값에서 물체의 화소값으로 바뀌게 된다. 이 점을 이용하기 위해서 시스템 수행 초반에 미리 배경 영상을 획득한다. 이후 영상을 획득할 때마다 배경 영상과의 차이를 계산한다.

$$f(x,y) = \begin{cases} f(x,y), (|R_1 - R_2| + |G_1 - G_2| + |B_1 - B_2|) > K \\ 0, \text{ else.} \end{cases} \quad (2)$$

$f(x,y)$: Original Image Valueable (x,y)
 R_1, G_1, B_1 : Background Image Pixel Data
 R_2, G_2, B_2 : Current Image Pixel Data
 K : Constant

식 (2)와 같이 차이 계산은 획득한 영상과 배경 영상의 동일한 위치의 화소의 R, G, B 값끼리의 차이를 절대값을 구하여 합하고 이 값이 일정 문턱 값보다 크면 획득한 영상의 화소값을 가지게 하고, 작으면 0의 값을 갖도록 한다. 이러한 과정을 거치고 나면 물체가 들어온 부분은 남게 되고 그렇지 않은 곳은 0의 값을 가지게 되어 검게 처리 된다.

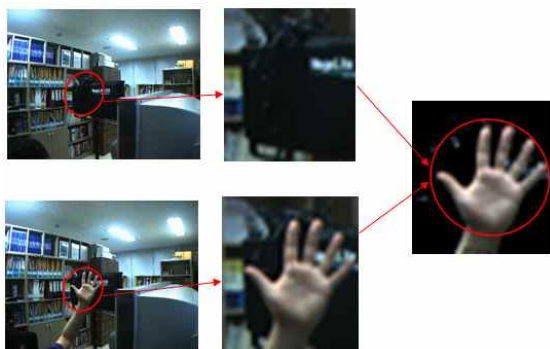


그림 3. 배경 영상 지우는 과정

배경이 제거된 영상에서는 물체만이 남게 된다. 그중에서 손 영역만을 분리하기 위해서 피부색 검출을 수행하고 손 영역 이외에 불필요한 부분을 제거한다. 사람마다 약간의 차이는 있지만 이를 전체 색차공간에서 보면 협소한 분포를 가지는 부분으로 구분할 수 있고 물체의

화소값을 이 분포의 포함 유무에 따라 피부영역과 피부영역 외의 부분을 구분한다.

피부색 검출에서 가장 민감한 부분은 조명에 의한 화소값의 변화이다. 조명의 위치에 따라서 피부색이 밝거나 어두워지게 되면 화소값은 정해놓은 분포를 벗어나게 되고 이는 판별 오류로 나타나게 된다. 따라서 조명 변화에 덜 민감한 YUV 색차공간을 사용한다. 획득한 영상은 RGB 색차공간이기 때문에 이를 식 (3)의 변환 공식을 통해 YUV 색차공간으로 변환한다.

$$\begin{aligned} Y &= (0.257 \times R + 0.504 \times G + 0.098 \times B + 16) \\ U &= (-0.148 \times R + 0.291 \times G + 0.439 \times B + 128) \\ V &= (0.439 \times R - 0.368 \times G - 0.071 \times B + 128). \end{aligned} \quad (3)$$

YUV 색차공간에서 피부색의 분포는 식 (4)와 같이 정의하였다. 이는 여러 번의 실험을 통해서 결정하였다.

$$\begin{aligned} 95 &\leq U \leq 135 \\ 115 &\leq V \leq 145. \end{aligned} \quad (4)$$

피부색을 판별하여 분포내의 값이면 획득 영상의 본래의 화소값을 가지고, 분포외의 값이면 0의 값을 부여하여, 피부색을 추출한 이미지를 생성한다.

계산을 쉽게 하기 위해 피부색 추출 영상을 사용하여 0 또는 255의 두 값만을 가지는 이진화 영상을 생성한다. 피부색 추출 영상의 화소의 녹색(green) 성분을 이용하여 green 값이 일정 문턱 값보다 크면 255를 작으면 0의 값을 부여하여 이진화 영상을 만든다. 그림 4에 피부색으로 검출된 영상과 이진화된 영상을 보였다.



그림 4. 색차정보의 특징을 이용하여 피부색 검출과 이진화한 영상
(a) 피부색 검출 영상 (b) 이진화 영상

생성된 이진화 영상에 손과 잡음을 또는 손과 얼굴을 구별하기 위해 서로 밀접된 화소에 같은 값을 부여하여 서로 구분할 수 있는 레이블링(labeling)을 수행한다. 레이블링의 동작은 영상 전체를 두 번 스캔하여 이루어진다. 스캔은 좌측 상단부터 우측 하단 방향으로 수행한다. 첫 스캔은 첫 탐색되는 화소부터 작은 값을 부여한다. 화소가 탐색되었을 때 좌측 위를 확인하여 번호를 부여 받은 화소가 있다면 같은 값을 부여하고 없는 경우는 새로운 값을 부여한다. 두 번째 스캔의 기능은 인접 하지만 다른 값을 가진 화소의 값을 두 값 중 작은 값을 부여하기 위함이다. 레이블링을 통해 얻어진 영상은 손과 나머지 영상의 분리도 가능하지만 같은 값들끼리 화소 수를 세어 손이라 판단하기 어려운 너무 많은 화소를 가진 값과 너무 적은 수를 가진 값들을 제거하는 기능도 할 수 있다. 그림 5는 레이블링의 수행 과정을 보여 주고 있다.

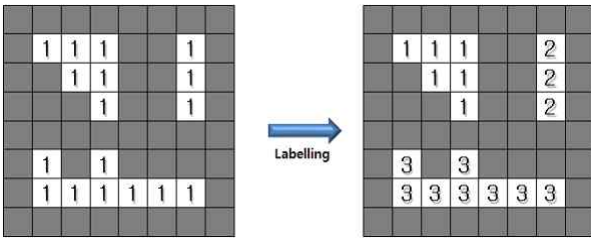
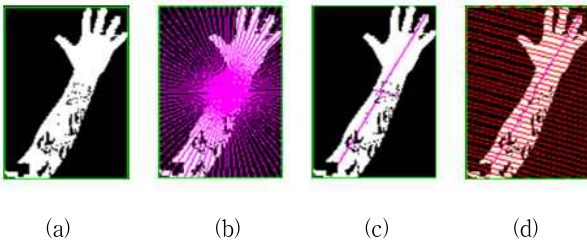


그림 5. 레이블링을 통한 영상의 분할

2-2. 손의 특징 점 찾는 방법

제안한 시스템은 사람의 정면 방향을 찍는 방식이기 때문에 사람의 팔꿈치 이전까지 영상이 찍힐 가능성이 높다. 따라서 손목 아래의 팔 부분을 제거하여 손의 동작을 판별하기 위한 손바닥과 손가락만의 영역을 획득해야 한다. 본 논문에서 제안하는 기법에서는 손목을 이용하여 손 영역을 분리한다.

그림 6의 (a)와 같이 레이블이 된 영상이 획득되면 손 영역을 포함하는 사각 영역을 설정한다. 영상의 전체 영역을 탐색하면서 손 영역의 화소를 지날 때의 X 좌표와 Y 좌표를 계산하고 이들 중 X 좌표와 Y 좌표의 최대값과 최소값을 판별한다. 이 4개의 값을 조합하여 4개의 꼭지점을 만든다. 꼭지점을 통해 만든 사각 영역을 손 영역으로 설정하고 영역 내에서만 연산을 하여 수행시간을 줄인다.



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

그림 6. 손 영역 찾는 법

(a) 손 영역 사각 영상 (b) 손의 길이 측정 영상

(c) 길이 최소, 최대값 찾은 영상 (d) 손 영역 내의 길이 찾은 영상

이렇게 만들어진 사각 영역의 중심점을 구하고 중심점에서 0~180°까지 선을 그려 선과 겹치는 손 영역의 양 끝점을 구해서 거리를 계산한다. 거리를 계산하는 방법으로 선과 겹치는 화소의 수를 세는 방법과 양 끝점의 X좌표, Y좌표 값을 이용하여 계산하는 방법을 제안한다.

손 영역을 추출할 때 오류로 인해 손 영역 안에 잡음으로 인한 홀 영역이 생길 수 있다. 두 방법 중 화소수를 세는 방법은 실제로는 거리가 더 길지만 홀 영역을 지날 때 화소를 세지 못하여 잘못된 거리가 계산될 확률이 높다. 따라서 양 끝점을 구해 계산하는 방법이 더 정확성이 높다. X 좌표와 Y 좌표를 이용하여 계산된 유클리안 거리를 식 (5)을 이용하여 구한다.

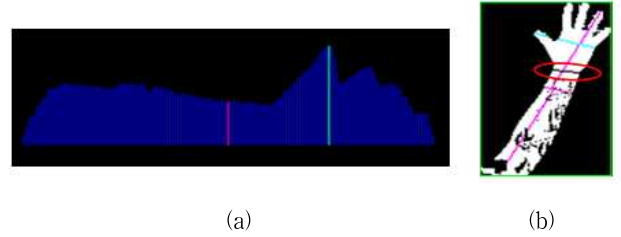
$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (5)$$

D : Distance
 x_1, y_1 : Pixel 1/x,y coordinates
 x_2, y_2 : Pixel 2/x,y coordinates

이때 각도 별로 계산한 값들 중 최대 거리는 그림 6의 (c)와 같이 팔의 끝부터 손가락까지의 거리일 가능성이 높다. 따라서 이때의 각도

는 팔과 손의 대략적인 각도를 알 수 있는 값이다.

그림 6의 (d)와 같이 최대 거리의 선의 화소들을 중심으로 하고 최고거리 일 때의 각도로 선을 그려 양 끝점을 구하고 유클리안 거리를 계산한다. 이 값들을 위치에 따라 전부 저장하고 팔부터 손까지의 굵기를 알 수 있는 손 영역히스토그램을 그림 7과 같이 생성한다.



(a) (b)

그림 7. 손 영역 히스토그램과 손목 찾기

(a) 손 영역 히스토그램 (b) 손목 찾은 영상

생성된 손 영역의 히스토그램의 값들을 비교하여 최대값을 구한다. 이때의 최대값은 엄지와 손바닥을 지나는 거리일 가능성이 높기 때문에 손목을 찾았을 때의 손과 제거될 팔의 방향을 구별 할 때 사용한다. 또 사각 영역의 중심과 최대값일 때의 위치의 거리의 1.5배 만큼 히스토그램을 탐색하여 이중 최소값을 손목의 위치로 판별한다. 손목의 위치를 결정하고 나면 히스토그램의 최대값이 있는 위치를 손의 위치로 판별하고 반대방향을 팔의 위치로 판별 손 영역만을 추출 할 수 있다. 이렇게 추출된 손 영역 이미지에서 중심점을 식 (6)을 이용하여 계산하여 이를 손의 2차원 좌표로 결정한다. 손 영역의 모든 화소의 X 좌표와 Y 좌표를 더하고 이를 화소 수로 나누어 평균적인 값을 구한다. 이렇게 구한 중심 좌표는 빠른 수행속도와 비교적 정확한 값을 획득한다.

$$C(x, y) = \frac{\sum_{m=0}^N P_m(x, y)}{N} \quad (6)$$

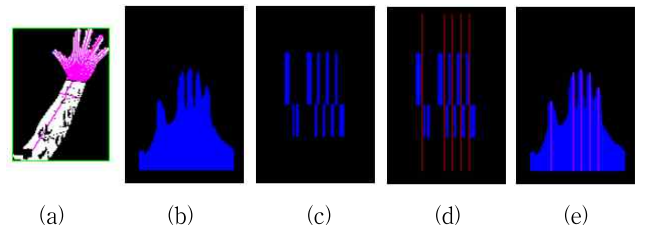
$C(x, y)$: 2D central coordinates

$P_m(x, y)$: summation of all coordinates of pixels decided skin

N : number of all pixels decided skin

2-3. 손 영역 히스토그램

손동작을 판별하기 위해 손목의 중심으로부터 손바닥 방향으로 0~180°까지 선을 그려 손목의 중심점과 만나는 끝점의 거리를 계산한다. 이 값들을 각도에 따라 저장하여 손가락 히스토그램을 생성한다.



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

그림 8. 손 영역 거리 히스토그램

(a) 손목에서 손끝까지의 거리 측정 (b) 거리 히스토그램

(c) 손가락 부분 검색 (d) 손가락 검출

그림 8과 같이 얻어진 히스토그램은 사람의 손과 거의 흡사한 모양을 가지고 있기 때문에 각도와 거리를 통해서 특정 손가락의 유무를

판별 할 수 있다. 손 영역 히스토그램은 기울기 정보를 이용하여 손가락의 개수를 계산한다. 작은 오차도 기울기에 영향을 끼치기 때문에 필터를 사용해 오차를 제거하여 기울기 값들을 구해 기울기 히스토그램을 만든다. 양의 값과 음의 값을 이용하여 손가락을 지날 때는 기울기가 줄어드는 형태를 이용해 양의 기울기에서 음의 기울기로 변화 되는 지점을 손가락으로 판별한다.

2-4. 3차원 증강현실 시스템

증강현실은 사용자가 눈으로 보는 현실세계와 부가정보를 갖는 가상세계를 합쳐 하나의 영상으로 보여주는 가상현실 중 하나로서 현실 환경과 가상환경을 융합하는 복합형 가상현실 시스템이다.

현실세계를 가상세계로 보완해주는 개념인 증강현실은 컴퓨터 그래픽으로 만들어진 가상환경을 사용하지만 주연은 현실 환경이다. 컴퓨터그래픽은 현실 환경에서 필요한 정보를 제공하는 역할을 한다. 사용자가 보고 있는 실사 영상에 3차원 가상 영상을 겹침으로써 현실 환경과 가상화면의 구분이 모호해지도록 하는 것이다. 그림 9에 다양한 증강현실의 예를 보였다.



그림 9. 다양한 증강현실 예제

가상현실 기술은 가상환경에 사용자를 몰입하게 하여 실제 환경을 볼 수 없다. 하지만 실제 환경과 가상의 객체가 혼합된 증강현실기술은 사용자가 실제 환경을 볼 수 있게 하여 보다 나은 현실감을 제공한다.

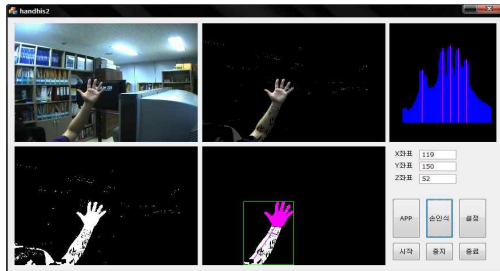


그림 10. 완성된 플레이어

그림 10은 완성된 플레이어로 버튼을 사용해 카메라의 시작, 중지, 설정을 할 수 있다. 또 손동작 인식 버튼을 이용해 손을 인식하기 시작하고 그에 따른 손 영역 히스토그램을 보여주고 계산된 3차원 좌표를 알려주게 된다. APP버튼을 클릭하면 손을 따라 3차원 객체가 표현되기 시작한다.

3. 실험 조건 및 결과

본 논문에서 제안한 기법을 사용하여 실시간 손 영역 인식에 대한 성능을 검증하기 위해 카메라에서 보이는 손의 각도와 손동작, 즉 손가락의 개수에 따라 손의 인식률을 측정하는 실험을 진행하였다. 표 1은 카메라에서의 손의 각도에 따라 손 인식률이 얼마나 되는지를 보여준

다.

표 1. 카메라와 손의 각도에 따른 인식률

	Recognition Rate
0°	86.2%
90°	75.8%
180°	84.6%
270°	87.0%

표 1에서 보여주는 결과를 통하여 스테레오 카메라를 통하여 얻은 결과는 평균 83%정도 나옴을 알 수 있었다.

표 2. 손가락 개수에 의한 인식률

Recognition Rate	85%	80%	84%	78%	76%

4. 결론

본 논문에서는 스테레오 카메라로부터 획득한 좌, 우 영상을 기반으로 피부색이 가지는 색차 정보의 특징을 이용하여 관심 영역인 손 영역을 추출하고 각각의 중심좌표를 구하여 증강현실 시스템을 구현하였다. 구현된 시스템에서는 중심좌표에 깊이 정보를 더하여 3차원 좌표를 완성하고 그 움직임으로 조작 가능한 가상공간을 제공하고 있다. 제안한 손 영역 인식 알고리즘에서는 관심 영역을 정의함으로써 계산 시간을 상당히 단축하였지만 여전히 높은 인식률을 보여 주고 있다.

5. 참고문헌

- [1] 이동석, 이동욱, 김수동, 김태중, 유지상 “스테레오 카메라를 이용한 이동객체의 실시간 추적과 거리 측정 시스템” 방송공학회논문지 제14권 제3호, 2009. 5, pp. 366 ~ 377
- [2] 황문구, 김형래, 강석범, 양태규 “퍼지 추론을 이용한 비전 기반 실시간 손모양 인식”. 한국정보기술학회논문지 제6권 제2호, 2008. 4, pp. 53 ~ 59
- [3] 김장운, 김승국, 장한별, 배기태, 이칠우 “에지 방향성 히스토그램 데이터를 이용한 손 형상 인식” 전남대학교 컴퓨터정보통신공학과. 한국정보과학회 2006 가을 학술발표논문집 제33권 제2호(B), 2006. 10, pp. 49 ~ 53
- [4] 복창순, 손연미, 방영철, 나보균 “손 인식을 이용한 컴퓨터 제어”. 한국정보과학회 2008 가을 학술발표논문집 제35권 제2호(B), 2008. 10, pp. 192 ~ 195
- [5] 박성욱, 방종관, 박중욱 “컬러 정보와 피부색 모델을 이용한 피부 영역 검출”. 전자공학회논문지-IE 電子工學會論文誌 第45卷 IE編 第4號, 2008. 12, pp. 60 ~ 67

감사의 글

- 본 연구는 지식경제부 및 정보통신연구진흥원의 IT산업원천기술개발사업의 일환으로 수행하였음. [2009-F-208-01, 대화형 디지털 홀로그램 통합서비스 시스템의 구현을 위한 신호처리 요소 기술 및 SoC 개발]