

# 스테레오 카메라를 이용한 3차원 공간 인식 시스템

\*이동석 \*김수동 \*이동욱 \*유지상

광운대학교

\*dlehdtr86@kw.ac.kr

## 3D world space recognition system using stereo camera

\*Lee, Dong-Seok \*Kim, Su-Dong \*Lee, Dong-Wook \*Yoo, Jisang

Kwangwoon University

### 요약

본 논문에서는 스테레오 카메라로부터 획득된 좌, 우 영상의 변이를 추정하여 3차원 공간 좌표(x, y, z)를 얻어내고, 거리 측정과 가상공간 제어를 통해 사용자에게 현실감을 제공하는 실시간 3차원 공간 인식 시스템을 제안한다. 스테레오 카메라로부터 획득된 좌, 우 영상은 시점의 차이 때문에 동일 물체에 대한 좌, 우 영상의 좌표 값의 차이를 발생시키는 테이를 변이(disparity)라 정의한다. 관심 영역의 변이를 추정할 때 일반적으로 관심 영역의 모든 화소(pixel)의 변이를 추정하지만, 제안한 알고리즘에서는 관심 영역의 2차원 중심 좌표(x, y)의 변이만을 추정하여 계산량을 줄이고 실시간 처리가 가능하도록 하였다. 카메라 파라미터를 이용하여 획득된 변이로부터 깊이 정보(depth)를 얻어내고 3차원 공간 좌표를 획득한다. 손을 관심 영역으로 설정한 시스템에서 3차원 공간 좌표는 실시간으로 사용자의 손의 움직임에 의해 획득되고, 가상공간(virtual space)에 적용되어 사용자가 가상공간을 조작할 수 있는 듯한 느낌을 준다. 실험을 통해 제안한 알고리즘이 1.5m 거리 내에서의 깊이 측정 시 평균 0.68cm의 오차를 가짐을 확인 할 수 있었다.

## 1. 서론

영상에서 깊이 정보를 얻는 방법에 대하여 많은 연구가 진행되어 왔다. 그 중에서 3차원 공간상에 설치된 스테레오 카메라로부터 얻어낸 좌, 우 영상에서 상호간의 정합점을 찾아내고 정합점간의 차이를 이용하여 3차원 깊이 정보를 추출하는 스테레오 비전(stereo vision)은 인간의 시각 시스템을 컴퓨터로 구현하려는 컴퓨터 비전(computer vision)분야에 있어서도 매우 중요한 부분으로 많은 연구가 진행되고 있다. 그러나 대응점을 추출하는 정합과정은 정밀성이 요구될 뿐만 아니라 연산에 상당한 시간이 소요된다[1].

이에 본 연구에서는 정합과정에 소요되는 시간을 줄여 실시간 처리가 가능하도록 좌, 우 영상의 관심 영역을 대표하는 특징점을 얻어내고 특징점만을 비교하여 변이 추정(disparity estimation)에 필요한 비교 영역을 줄여 속도를 향상시키는 알고리즘을 제안한다. 또한 관심 영역의 움직임을 3차원 좌표로 추적하여 사용자에게 제공한다. 깊이 정보를 포함한 3차원 움직임을 가상공간을 통하여 사용자에게 실시간으로 보여주어 직접 조작하는 것 같은 현실감을 준다. 실험을 통하여 추출된 깊이 정보와 실제 거리와의 비교를 통하여 제안한 알고리즘을 통한 변이 추정의 정확성을 입증하였다.

본 논문의 구성은 다음과 같다. 2장에서는 전체적인 구조도와 세부적인 동작을 설명하고, 3장에서는 제안한 변이 추정 알고리즘의 정밀성을 측정하는 실험을 수행한다.

## 2. 시스템 구조도

그림 1에서는 시스템의 전체적인 흐름을 보여준다. 스테레오 카메라로부터 얻은 좌, 우 영상에서 피부색의 특징을 이용하여 관심 영역을 얻어낸다. 관심 영역의 특징점을 구하여 실시간으로 변이를 추정하고 깊이 정보를 포함한 3차원 좌표를 완성하여 가상공간과 연결한다.

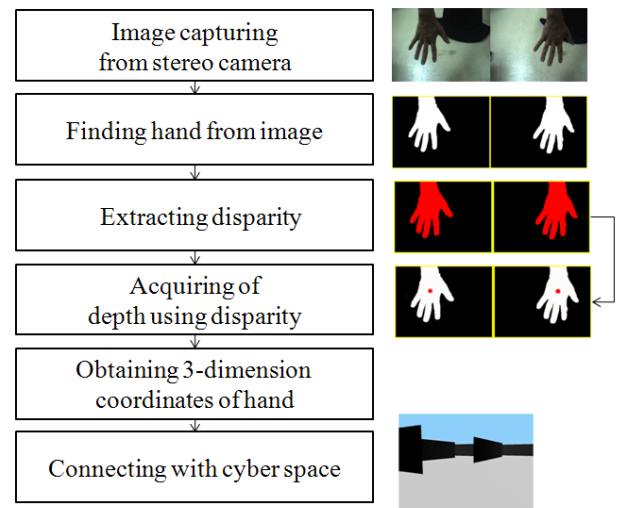


그림 1. 제안한 시스템 구조도

### 2-1. 관심 영역 분리

변이 추정을 하기 위한 여러 가지 방법 중에서 제안한 시스템에서는 관심 영역만 따로 분리하여 변이를 추정하는 방법을 사용한다. 이를

위해서는 관심 영역이 비관심 영역과 정확히 구분 되어야 한다. 본 논문에서는 관심 영역 분리를 위해서 그림 2와 같은 과정을 거친다.

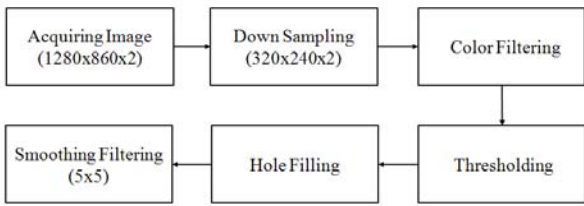


그림 2 . 관심 영역 분리하는 흐름도

카메라에서 제공된 한 프레임(frame)은 1280x960의 크기를 가진다. 이는 크기에 따른 연산 량이 많아서 실시간(초당 15 frames 이상)으로 처리하기에 부적합하다. 연산량을 줄이기 위해 영상을 가로, 세로 각각 4배로 down sampling을 수행하여 크기를 320x240으로 줄여 연산해야 할 정보량을 1/16로 줄였다. 변환된 영상에서 관심 영역을 추출하고 변이를 추정한다. 제안한 시스템은 사용자가 쉽게 다룰 수 있도록 손을 관심 영역으로 정하였다. 사람의 피부색 분포는 색차 평면상에서 자연계의 다른 물체와 구분할 수 있는 식(1)과 같은 특징을 가지고 있다.

$$\begin{aligned}
 30 &\leq Y \leq 130 \\
 100 &\leq U \leq 129 \\
 128 &\leq V \leq 137
 \end{aligned} \tag{1}$$

사람마다 다소 차이는 존재하지만 색차 평면상의 피부색의 분포는 위의 식(1)과 같이 매우 협소한 영역을 차지하고 있다. 또한 YUV 구조가 RGB 구조보다 조명 값의 변화에 덜 민감하기 때문에 아래의 식(2)를 이용하여 RGB구조를 YUV구조로 변환하여 피부색 영역을 구분한다[4].

$$\begin{aligned}
 Y &= (0.257 \times R + 0.504 \times G + 0.098 \times B + 16) \\
 U &= (-0.148 \times R - 0.291 \times G + 0.439 \times B + 128) \\
 V &= (0.439 \times R - 0.368 \times G - 0.071 \times B + 128)
 \end{aligned} \tag{2}$$

변환된 좌, 우 영상에 식(1)의 조건을 이용하여 이진화시킨 영상이 그림 3과 같다.

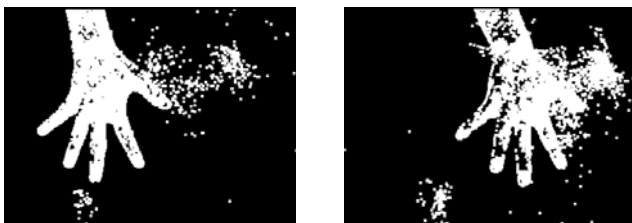


그림 3 . 색차정보의 특징을 이용하여 이진화한 영상  
(a) 좌 영상 (b) 우 영상

그림 3의 결과는 카메라가 조명의 방향이나 강도에 따라 관심 영역에 잡음(noise)이 발생한다는 것을 보여준다. 잡음을 제거하기 위해 hole filling과 smoothing filtering 두 가지 과정을 수행한다.

hole filling은 영상의 심한 잡음을 제거하는 데 사용한다. 화소 단위로 영상을 탐색하면서 처음 화소와 같은 값을 가지는 화소 사이의 좌표 값의 차이가 일정 값 보다 작으면 그 사이의 모든 화소의 값을 처음 화소의 값으로 변경시킨다. 반대로 같은 값을 가지는 화소끼리의 좌표 차이가 지정한 값보다 크다면 사이의 모든 화소 값을 변경시키지 않는다. hole filling은 정렬 과정이 없어 빠른 속도로 수행되고 수직과 수평, 두 방향으로 수행한다. hole filling은 관심 영역 내의 잡음은 효과적으로 제거하지만 그림 3과 같이 관심 영역 이외의 부분에서 생성된 잡음은 제거하지 못하고 잡음을 더하게 되는 단점을 가지고 있다. hole filling으로 제거하지 못한 관심 영역 외의 잡음을 제거하기 위해 smoothing filtering을 수행한다. 화소 단위로 영상을 탐색하면서 해당 화소를 중심으로 하는 마스크 내의 0의 값을 가지는 화소의 개수를 세어 일정 개수를 넘어가면 해당 화소의 값을 0으로 변경시키는 동작을 수행한다. smoothing filter는 저역통과필터(lowpass filter)를 수행한 것처럼 영상이 다소 흐려지지만 hole filling이 제거하지 못하는 관심 영역 외의 심한 잡음을 제거하는 데 효과적이다. smoothing filter는 마스크의 크기에 따라 잡음 제거의 효과는 증가하지만 속도가 느려지는 단점이 존재한다. 본 논문에서는 여러 크기의 마스크로 실험해 본 결과 5x5 크기의 마스크가 가장 적합하였다. 아래의 그림 4는 hole filling과 smoothing filtering을 차례로 거친 영상의 결과를 보여준다.



그림 4 . hole filling과 smoothing filtering을 수행한 후의 영상  
(a) 좌 영상 (b) 우 영상

그림 4의 결과 영상은 hole filling과 smoothing filtering이 정확하게 관심 영역을 추출하는 데 있어서 효과적이라는 것을 확인 할 수 있다.

## 2-2. 변이 추정 기법

다양한 변이 추정 기법 중에서 본 연구에서는 일정 영역의 영상조각(image patch)을 가지고 다른 영상과 상관관계를 이용하여 변이를 추정하는 영역기반(area-based)정합을 사용한다. 영역기반정합은 주로 마스크 단위로 비교하여 변이를 추정하는 형태이기 때문에 마스크의 크기에 따라 결과가 유동적이다. 마스크의 크기가 작으면 속도가 빠르지만 잡음에 취약하고 마스크의 크기가 크면 잡음에 강하지만 속도가 느리고 일종의 저역통과필터의 역할을 하여 세밀한 처리가 불가능하다. 이 같은 단점을 보완하기 위해 일반적으로 그림 5와 같은 영역 분할 기법을 사용하여 변이 추정에 있어 속도와 정확도를 높인다[1]. 그림 5와 같이 좌, 우 영상의 모든 관심 영역을 이용하여 변이를 추정하는 영역 분할 기법에 비해 본 논문에서는 좌, 우 관심 영역의 특징점을 추출하고 비교하는 방식으로 변이를 추정하여 빠른 속도의 변이 추정 기법을 제안한다. 관심 영역이 분리된 영상에서 관심 영역으로 판단

한 모든 화소의 좌표 값을 더하여 그 개수로 나누어 수직, 수평적으로 평균이 되는 2차원 중심좌표를 얻어낸다. 중심좌표를 얻는 식은 식(3)과 같다.

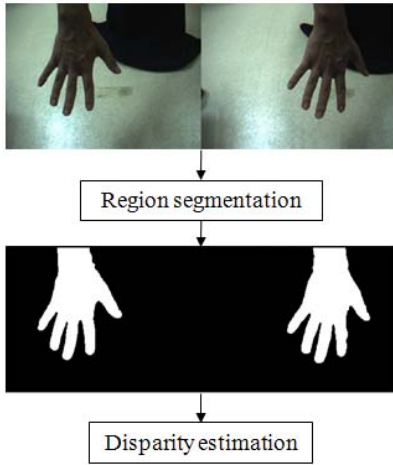


그림 5. 영역 분할 변이 추정

$$C(x,y) = \frac{\sum_{m=0}^N P_m(x,y)}{N} \quad (3)$$

$C(x,y)$  : 2D central coordinates  
 $P_m(x,y)$  : summation of all coordinates of pixels decided skin  
 $N$  : number of all pixels decided skin

그림 6은 중심좌표를 특징점으로 하여 변이를 추정하는 영역 분할 방식과 기존의 영역 분할 기법의 비교 범위의 차이를 보여준다. 제안한 변이 추정 방식은 변이를 추정하기 위해 비교하는 범위가 현저하게 줄어들어서 변이 추정하는 속도가 향상되었다.

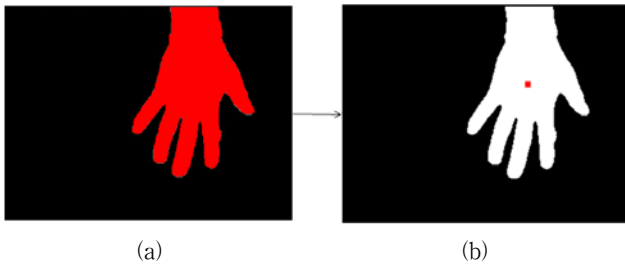


그림 6. 변이 추정을 위한 비교 범위의 축소  
 (a) 기존의 방법 (b) 제안한 방법

좌, 우 영상의 2차원 중심좌표를 아래의 식(4)의 식에 대입하여 깊이 정보를 얻어낸다. 식(4)에서  $Z$ 는 계산된 실제 거리를 뜻한다. 식(3)을 이용하여 얻은 2차원 중심좌표에 깊이 정보를 추가하여 3차원 좌표를 완성한다.  $B$ (baseline)은 카메라의 두 렌즈의 중심(optical center) 간의 거리를 나타내고  $f$ (camera focal length)는 두 렌즈의 초점거리를 나타낸다.  $C_L, C_R$ 은 좌, 우 영상의 2차원 중심좌표이다.  $a$ 는 실험을 통해 계산하여야 하는 임의의 값으로 원하는 단위의 값으로 환산할 때 사용하는 값이다.

$$Z = \alpha f \frac{B}{C_R - C_L} \quad (4)$$

$Z$  : real distance  
 $\alpha$  : scale factor  
 $f$  : camera focal length  
 $B$  : baseline  
 $C_R$  : central coordinate of right image  
 $C_L$  : central coordinate of left image

### 2-3. 가상공간의 구현

3차원 좌표가 마우스 포인터를 추적하는 동작을 수행하는 플레이어를 제작하였다. 또한 마우스 포인터의 움직임으로 제어하는 가상공간을 제작하여 플레이어와 연계시켜 주었다. 3차원 가상공간은 OPENGL 이라는 API(application programming interface)를 사용하여 제작하였다. 다음 그림 7은 제작한 가상공간을 보여준다.

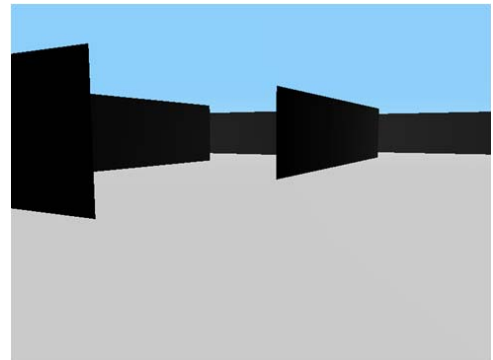


그림 7. 가상공간

가상공간은 마우스 포인터의 움직임을 따라 미로 찾기를 할 수 있도록 제작하였다. 관심 영역의 상하, 좌우, 전후의 모든 움직임을 따르도록 제작하여 사용자는 관심 영역의 움직임으로 가상공간 상의 위치를 조정할 수 있도록 하였다.

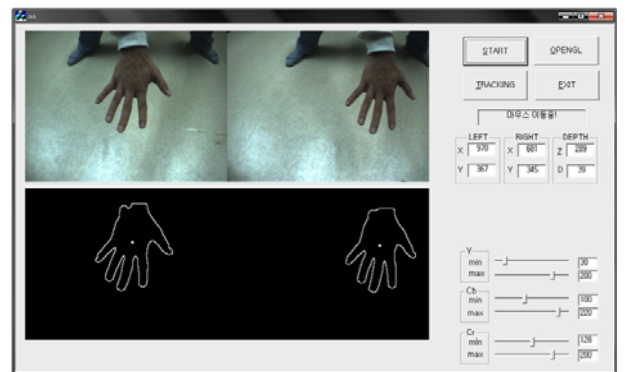


그림 8. 완성된 플레이어

그림 8은 모든 알고리즘과 동작을 포함하여 제작한 플레이어를 보여준다. 버튼을 통해 카메라로부터 얻은 좌, 우 원 영상과 알고리즘을 거쳐 변이 추정에 쓰이는 영상을 보여 주고 3차원 좌표를 마우스 포인터와 연결시키는 추적(tracking)동작과 OPENGL을 이용하여 제작한 가상공간은 별도의 버튼으로 동작 시키도록 하였다. 간단한 마우스의 동작을 설명하는 박스와 좌, 우 영상의 2차원 중심좌표, 그리고 알고리

즘을 통해 추정된 특징점 간의 차이와 식(4)를 이용하여 얻은 실제 거리의 측정치를 실시간으로 보여주도록 제작하였다. 조명에 영향을 비교적 덜 받는 YUV 색차 구조를 사용하여 피부를 추출하지만 사용 환경에 따라 피부로 인식되는 색차 정보의 값도 유동적이기 때문에 사용자가 색차 정보를 조절할 수 있도록 컨트롤러로 구현하였다.

### 3. 실험 조건 및 결과

본 논문에서 제안한 알고리즘을 사용하여 실시간 변이 추정의 성능을 검증하기 위해 거리에 따라 얻은 깊이 정보와 실제 거리 값과의 오차를 측정하는 실험을 진행하였다. 그림9는 카메라로부터의 거리에 따른 좌, 우 영상의 특징점간의 좌표의 차이를 보여준다.

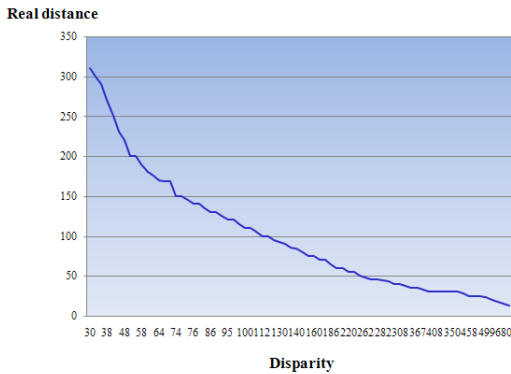


그림 9 . 거리에 따라 측정된 변이

그림 9의 실험 결과를 통하여 센티미터(cm)단위로 하였을 때 식(4)의  $\alpha$ 는 15.74의 값을 얻어 내었다. 실험에 사용한 스테레오 카메라(모델명 : BUMBLEBEE XB3)의 B값(12)과 F값(60)을 식(4)에 대입하여 깊이 정보를 획득하였다. 식(5)와 식(6)을 통해 획득한 깊이 정보와 실제 거리의 오차와 평균 오차거리를 측정하였다.

$$\text{오차(cm)} = | \text{실제 거리} - \text{측정 거리} | \quad (5)$$

$$\text{평균 오차거리(cm)} = \text{시행한 오차의 총 합} / \text{실험 시행 횟수} \quad (6)$$

표 1 . 측정거리의 평균 오차

카메라로부터의 거리(cm)	평균 오차거리(cm)
20 ~ 40	1.75
40 ~ 150	0.5
150 ~ 250	6.8

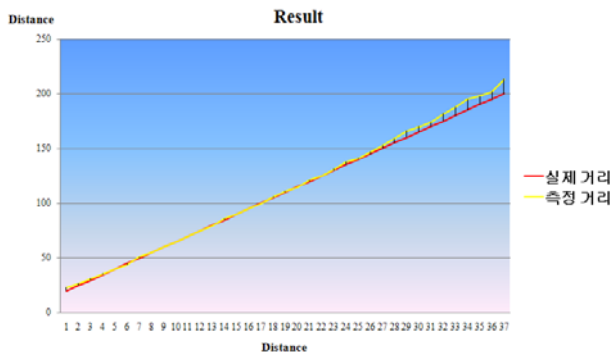


그림 10 . 거리에 따른 측정거리와 실제거리의 오차

표(1)과 그림(10)은 카메라로부터의 거리에 따른 오차를 보여준다. 실험 결과 근거리(20~150cm)에서는 비교적 높은 신뢰도를 보였지만 거리가 멀어져 200cm를 넘어가면서 부터 선형적으로 오차가 커지는 것을 확인 하였다. 실시간 처리를 위해 제한한 변이 추정 알고리즘이 근거리에서 비교적 정확한 변이를 추정하는 것을 확인 할 수 있었다.

### 4. 결론

본 논문에서 구현한 시스템에서는 스테레오 카메라로부터 획득한 좌, 우 영상을 기반으로 피부색이 가지는 색차 정보의 특징을 이용하여 관심 영역을 획득하고 각각의 중심좌표를 구하여 변이를 추정하여 깊이 정보를 추출한다. 중심좌표에 깊이 정보를 더하여 3차원 좌표를 완성하고 그 움직임으로 조작 가능한 가상공간을 제공한다. 제안한 특징점 기반의 변이 추정 방법은 빠른 속도를 위해 비교 영역을 상당히 축소하였지만 근거리에서 높은 신뢰도를 보였다.

### 5. 참고문헌

- [1] 박상백, 문종섭, “동적계획법을 적용한 효과적인 폐색영역제거를 위한 고속 스테레오 정합법”, 2002.
- [2] 김계영, 장석우, “유형기반 스테레오 정합을 통한 영상 변이 측정”, 한국컴퓨터정보학회 논문지, 2006.
- [3] 김장운, 김송국, 장한별, 배기태, 이철우, “에지 방향성 히스토그램 데이터를 이용한 손 형상 인식”, 한국 정보과학회 학술발표논문집, 2006.
- [4] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Second Edition.

감사의 글

본 연구는 서울시 산학연 협력사업(NT080528) 지원으로 수행되었음.