

# 스테레오 비전 알고리즘을 이용한 향상된 트래킹 연산

이기정\*, 조형진\*, 송민규\*\*, 이병국\*\*\*

\*동아대학교 컴퓨터공학과, \*\*동서대학교 컴퓨터공학과

\*\*\*동서대학교 멀티미디어공학과

e-mail : [keycore79@hanmail.net](mailto:keycore79@hanmail.net)

## Advanced Tracking Calculation by Stereo Vision Algorithm

Ki-Jeong Lee\*, Hyung-Jin Cho\*, Min-Gyu Song\*\*, Byung-Gook Lee\*\*\*

\*Dept. of Computer Science, Dong-A University

\*\*Dept. of Computer Science, Dongseo University

\*\*\*Divison of Internet Engineering, Dongseo University

### 요 약

모션 캡처 기술은 영화나 게임 혹은 애니메이션과 같은 다양한 분야에서 새롭게 시도되고 있다. 기존의 모션 캡처 방식은 마커라는 불편한 인식장치로 인하여 움직임에 적지 않은 제약을 받을 뿐만 아니라 엄청난 고가 장비를 필요로 한다. 갈수록 좀 더 편리한 모션 캡처 기술이 요구됨에 따라 실시간으로 물체의 위치를 트래킹(위치추적)하고, 스테레오 비전을 이용하여 3 차원 재 구축을 수행해 입체적인 가상모델을 생성하여 보았다. 본 논문에서는 효율적인 트래킹 연산에 의한 움직임 제약을 최소화한 개선된 알고리즘을 설계, 구현하였고, 저렴한 웹 캠을 이용하여 스테레오 비전방식(Stereo Vision Based)을 접목시켜 기존의 고가 장비들과 유사한 환경을 이 시스템으로도 가능하다는 것을 확인하였다.

### 1. 서론

인간이나 물체의 행동은 움직임을 통해 의미를 전달한다. 이러한 의미 전달기술은 컴퓨터 산업의 비약적인 발전으로 인하여 다양한 멀티미디어 분야에서 시도하고 있다.

본 논문에서는 실 세계에 존재하는 물체를 시간과 공간 데이터로 변환하여 두 대의 카메라를 통해 입력 받은 물체의 이미지 정보를 3차원의 물체로 복원하여 실제 인간이 느끼는 깊이 정보를 컴퓨터에도 인지 시킬 수 있도록 하였다. 이러한 연구를 통해 물체 혹은 사람의 모션을 마커(marker)가 없이 실시간으로 캡처(capture)하여 가상현실에서 3차원 재 구축을 수행해 입체적인 가상 모델을 생성하여 성능 측정을 해보았다.

AR(Augmented Reality)에서 Real World와 가상의

오브젝트를 정확하게 정렬시키기 위해서 Real World에 존재하는 대상체의 움직임을 감지하여야 하는데 이를 수행하는 것을 Tracker라고 하며, 이때 감지하여야 할 물리량은 회전각과 위치정보이다. 트래킹을 수행하는 기술로는

- ① Mechanical Tracker
- ② Magnetic Tracker
- ③ Optical Tracker
- ④ Acoustic (Ultrasonic) Tracker
- ⑤ Inertial Tracker
- ⑥ Hybrid Tracker 등이 있다[8]

효율적으로 물체의 위치정보를 추출하기 위해서 개선된 트래킹 연산 알고리즘을 설계하여 저렴한 웹 캠을 스테레오 비전(Stereo Vision) 방식에 접목을 하여 고가의 장비들과 유사한 환경을 본 시스템

으로도 가능하다는 것을 확인하였다.

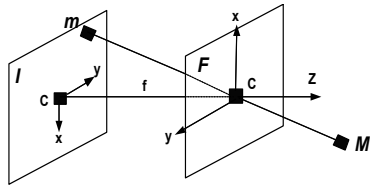
본 논문의 구성은 다음과 같다. 2장에서는 관련 기술에 대해서 기술하고, 3장에서는 트래킹 연산에 의한 3차원 스테레오 비전 시스템 구현 내역을 설명한다. 4장에서는 본 논문의 결론 및 향후 연구 방향에 대해 기술한다.

## 2. 관련기술

### 2.1 카메라와 영상

[그림 1]의 영상평면 I에서 초점거리 f의 장소에 I에 평행한 면 F를 놓고, 그 위의 점 C에 핀홀을 둔다. 물체에서 오는 빛은 핀홀(점C)를 통해 영상 평면에 상을 맺는다. 물체의 점과 핀홀, 영상 평면 상의 상은 동일한 일직선상에 있다.

이와 같은 사영을 중심사영(perspective projection)이라고 하며 점C를 렌즈중심(focal point), 또는 초점(focus), 면 F를 초점면(focal plane), 그리고 렌즈중심부터 영상평면까지의 거리 f를 초점거리(focal length)라고 한다. 점 C를 통해 영상평면에 수직인 선을 광축(optical axis)이라고 하고 그것과 영상평면과의 교점 C를 영상중심(principal point)이라고 한다. 또한 광축은 점면에도 직교한다.

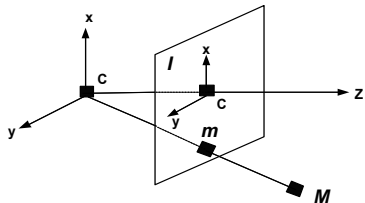


[그림 1] 핀홀 카메라 모델 (첫번째)

영상 좌표 계는 영상중심 C를 원점으로 하고 x축과 y축은 각각 카메라소자의 배치 축에 맞춘다. 광축을 제 3축으로 했을 때에 오른손 좌표계가 되도록 순서를 정한다. 새로이 3차원공간의 좌표계는 초점 C를 원점으로 광축을 Z축으로 하고 X축과 Y축은 각각 x축과 y축의 반대 방향에 둔다. 이와 같이 정의한 오른손 C-XYZ 좌표계를 카메라 좌표계(camera coordinate system)라고 부른다. 이상의 좌표계를 사용하면 3차원 공간 안의 한 점과 2차원상과의 사이에 다음의 관계가 성립한다.

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (1)$$

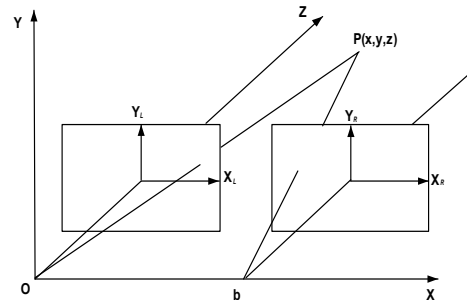
[그림 2]에서 보인 바와 같이 초점면 뒤에 있는 영상면을 초점면 앞으로 가져와 반전시켜도 상기의 관계가 성립한다. 컴퓨터 비전에서는 이 그림의 경우가 많이 이용되고 있다. [4]



[그림 2] 핀홀 카메라 모델 (두번째)

### 2.2 양안시(Stereo Vision)

스테레오 정합을 하기 위한 영상을 취득하는 카메라의 기하학적 구조를 잘 이용하면 스테레오 정합과정을 보다 쉽게 할 수 있다. 3차원의 물체 영상을 위치가 다른 두 지점에서 취득하기 때문에 각 영상에서 물체의 위치가 틀려져서 위치 변화의 양은 물체를 보는 각도에 따라 달라지게 된다. 특히 문제를 더 어렵게 하는 이유는 국부적인 점이나 면들이 다른 쪽의 영상에서 하나 이상의 정합 후보를 가지게 되므로 전체적인 일치성을 만족 하도록 국부적인 특징들을 서로서로 연계하여 후보를 고려해야 하기 때문이다. 일치성 문제를 해결할 때 정합할 점들에 대한 탐색 영역은 영상을 취득하는 카메라들 사이의 기하학적인 구조를 이용하면 상당히 줄어 들 수 있다. 일반적으로 두 카메라의 초점과 영상의 한 점을 지나는 평면을 epipolar line 이라고 한다. 한 점에 대한 정합 후보는 이 점에 대한 다른 쪽 영상에서의 epipolar line 위에만 존재하게 되므로 이 선상의 점들에 대해서만 정합을 시도하면 되는데 이를 epipolar 구속 조건이라고 한다. 특히 두 영상의 수평선이 일직선 상에 놓여있고 수직선이 평행한 경우 epipolar line은 각 영상의 수평선이 된다. 따라서 두 영상의 같은 라인상에 존재하는 특징들간의 정합만을 고려하여 탐색 영역을 줄일 수 있다.



[그림 3] 스테레오 비전

두 대 카메라의 초점거리 f가 서로 같고 좌측, 우측 카메라의 각 영상 좌표계를  $X_L - Y_L - Z_L$  과  $X_R - Y_R - Z_R$  이라고 할때, 두 카메라의 광축을 각각  $Z_L, Z_R$  로 하여 서로 평행하게 되도록 한다. 좌측 렌즈의 중심과 우측 렌즈의 중심사이의 거리를 b라고 하고 이 두 중심위치를 연결하는 선분을 베이스 라인(base line)이라 한다. 삼차원 공간 좌표계 X-Y-Z에 대해 좌측, 우측 영상 좌표계의 원점을 각각 (0, 0, f)와 (b, 0, f)에 놓고 좌측 영상의  $Z_L$  축과 공간 좌표계의 Z축을 서로 일치 시킨다.

Epipolar구속조건을 가정하면  $Y_L = Y_R$  즉, 다음 [그림3]과 같은 영상 취득 구조가 된다.

만일 3차원 공간상의 한점 P(x, y, z)가 좌, 우 영상면의  $(x_L, y_L)$ 과  $(x_R, y_R)$ 에 투영되었다고 가정하면 이들 사이의 관계는 삼각형의 비례관계를 이용하여 다음과 같은 식으로 나타낼 수 있다.

$$x_L = \frac{fx}{z} \quad (2)$$

$$x_R = \frac{f(x-b)}{z} \quad (3)$$

위의 두 식으로부터  $z$ 를 구하면,

$$z = \frac{bf}{x_L - x_R} = \frac{bf}{d} \quad (4)$$

이때  $d$ 를 변위(disparity)라고 하며 물리적 의미로는 인간시각에서 두 눈의 시차와 같은 의미를 갖는다. 위의 식에서  $b$ 와  $f$ 는 카메라와 영상취득 구조에 의해 결정되는 값이므로 두 영상의 변위를 구하면 3차원 깊이 정보인  $z$ 값을 구할 수 있다. 또한 이 식으로부터 알 수 있는 것은 깊이  $z$ 와 변위  $d$ 는 서로 반비례 관계가 있으므로 카메라와 가까이 위치한 물체일수록 보다 정확한 깊이 정보를 구할 수 있다. 즉, 멀리 위치한 물체 경우 변위  $d$ 의 값이 작게 되므로 변위의 작은 오차에도 깊이 정보는 크게 틀려질 수 있으나 가까이 위치한 물체 경우에는 변위 값이 크게 되므로 변위의 작은 오차에도 깊이 정보의 변화량은 상대적으로 크지 않으므로 보다 정확한 깊이를 구할 수 있다. 그리고 변위는  $b$ 에 비례하므로 변위를 구할 때 생기는 오차 값이 일정하게 고정되면 구하는 깊이 정보는  $b$ 값이 크면 클수록 보다 정확한 값을 구할 수 있게 된다. 그러나 두 카메라 사이의 거리가 멀어지면 멀어질수록 두 카메라에 투영되는 공동 가시영역(common view area)이 줄어들게 되어 카메라에 가까이 위치한 물체들은 공동가시 영역에서 벗어나게 되어 정합할 수 없는 문제점이 있다. 또한 변위는 초점 거리에 비례하고 초점거리에 따라 투영된 영상이 확대 또는 축소되므로 사용하는 카메라의 초점거리가 결정되면 깊이 정보를 구하고자 하는 물체들이 공동가시 영역에 모두 놓일 수 있도록 두 카메라 사이의 거리  $b$ 를 결정하여야 한다.[1][3]

## 2.3 영상처리를 이용한 색깔 추출

### 2.3.1 이차원 영상 처리(Blob feature analysis)

- ① RGB칼라 모델을 사용함으로써 차이값을 단편화
- ② YIQ 칼라 모델을 사용함으로써 색깔의 정점 지역을 추출
- ③ 룩업(Lookup Table)에 의한 YIQ연산의 빠른 계산적용[7]
- ④ 가장 적합한 YIQ 임계값을 적용하여 색깔 (빨강, 연두, 파랑, 피부색)을 추출

### 2.3.2 영상 처리의 연산속도 과부하(640X480 pixels)

한 프레임의 이미지(640X480)를 처리하기 위해선 debugging(1번)과 four colors(4번)의 과정을 거쳐야 한다. 이것은 CPU과부하와 함께 실시간 전체 5번 연산을 요구하여 하드웨어 spec의 한계와 집결된다. 하지만 본 논문에서 서술한 트래킹 알고리즘에서는

동적인 트래킹 알고리즘을 통하여 색깔을 찾아냈으며 이를 통해 연산률도 개선되고 인식률도 높아짐을 확인할 수 있었다.[표 1]

## 3. 트래킹 연산에 의한 3차원 비전 시스템

### 3.1 색 영역 추출, 색 영역 트래킹

#### 색 영역 추출

실 세계의 색깔 데이터를 컴퓨터에 저장하게 되면 아날로그 데이터로부터 디지털 데이터로 컴퓨터로 변환, 저장 되는데 이 때에 색깔 정보 값이 손실되어 저장이 된다. 따라서 RGB모델로는 색깔 찾기가 어렵다. 일반적으로 알려진 다음의 두 방식으로 변환을 해서 이미지 처리를 수행한다.

- ① YIQ model (by television broadcasting)
- ② HSV model (easy to recognize by human)

본 논문에서는 ①번 모델을 사용하여 색깔추출을 하였다.[5][6][7]

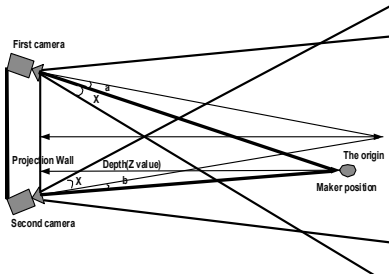
#### 색 영역 트래킹 방법

- ① 각각의 색깔에 대한 초기 영역 값 추출
- ② 추출된 각각의 정보를 저장
- ③ 다음 프레임 데이터가 들어올 때 포인트주변을 탐색하고 각각의 추출영역에 대해 트래킹을 수행
- ④ 트래킹 알고리즘을 사용해서 연산을 개선시킴

예를 들어 하나의 사각형이 10픽셀이라면 스캔 수행 시 단지 80\*80의 연산만 필요 된다. 이는 4배 이상의 성능 개선과 직결되었다.

### 3.2 삼각측량(Triangulation)

한 쌍의 중복된 사진으로부터 각 점의 3차원 절대 좌표를 측정하기 위해서는 최소한 2개의 평면 기준점과 3개의 표고 기준점이 요구된다. 이들 기준점을 획득하기 위해 필요한 모든 점을 측정하는 것을 전면 지상 기준점 측량(Full Ground Control Point Survey)이라고 하는데, 대규모의 항공 사진들을 이용하여 작업을 수행하는 경우 이러한 지상 기준점 측량 작업은 엄청난 시간과 노력, 비용의 소모를 가져온다. 따라서, 실제의 작업에서는 소수의 지상 기준점에 대해서만 측량을 실시하고 나머지 점들에 대해서는 측정된 지상 기준점의 좌표와, 도화기 등의 정밀 좌표 측정기에 얻어진 사진 좌표나 모델 좌표 또는 스트립 좌표 등을 이용하여 수학적 계산으로 절대 좌표를 결정하게 되는데 이러한 방식을 항공 삼각측량이라고 한다.[2][9]



[그림 4] 스테레오 비전 평면도

[그림 4]에서는 트래킹 공간을 수학 함수로써 표현하였고, 스테레오 비전환경을 위에서 본 평면도로써 보여준다. The Origin(원점)의 위치는 물체의 초기 위치 값이며 수학 함수를 통하여 마커의 이동을 좌표 값으로 얻어낸다.

그리고 각 카메라가 가질 수 있는 최대 영역 값을 고려해서 각도와 길이를 이용한 물체의 변위 값을 찾아가는 과정을 보여준다.[10]

### 3.3 트래킹 연산에 의한 개선된 알고리즘

다음은 트래킹을 하기위한 소스이다

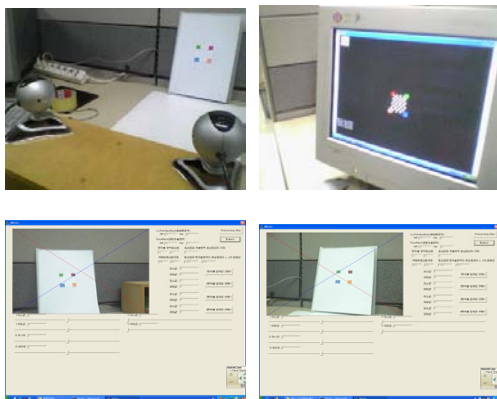
```

Client :
Boolean BinaryByYIQ(Y, I, Q) :: =
    return RGB 임계값을 적용하여 YIQ특업 테이블을 이용한
        고속 이진화 처리
CRect FindColor() :: =
    return 라벨링 알고리즘을 적용하여 색상 영역 추출
Size GetSize() :: =
    return 색상 영역 결정후 최적화 사이즈 추출
CPoint Tracking() :: =
    return 최적화 사이즈 크기를 적용하여 실시간 트래킹
        연산 후 중심점 추출

Server :
Boolean SteroMatching(Left, Right) :: =
    return 좌우 영상추출 색상영역의 epipolar line 탐색 후
        정합연산 처리
Int SteroVision() :: =
    return 삼각측량 기법을 이용하여 색상영역 공간 상의
        Z값 추출
Void Draw3D(X, Y, Z) :: =
    return 추출된 점(X, Y, Z)을 이용하여 3D 복원
    
```

[그림 5] 개선된 트래킹 알고리즘

### 3.4 전체 시스템 구성도



[그림 6] 구현된 시스템 구성도  
(<http://www.markerfree.ww.to>)

[그림 6]은 개선된 알고리즘을 적용하여 만든 시스템의 전체 구조이다. 좌상은 스테레오 비전을 세팅하여 촬영하는 장면이고 우상은 가상의 물체를 3D로 복원, 나머지 그림들은 스테레오 비전에서 좌, 우 각각의 카메라를 개선된 알고리즘으로 색깔을 추출하고 동시에 트래킹을 하고 있는 장면이다.

### 3.4 성능 측정 및 비교

개선된 트래킹 알고리즘 적용 시 초당 15 프레임의 성능을 보였다. 이는 개선된 트래킹 알고리즘을 사용하지 않았을 시 보여주었던 초당 2~3 프레임에 비하여 확실한 성능향상을 보여주었다.

[표 1] 개선된 알고리즘 적용 전, 후 비교

카메라	프레임	카메라 사양	(단위 : 프레임 / 초)	
			(적용 전)	(적용 후)
고사양(CCD)		300~1000	30~100	210~700
저사양(웹캠)		25~30	2~3	15~17

### 4. 결론

가능하다면 최신 CCD 카메라와 같은 고가의 카메라를 사용하면 영상의 왜곡도와 카메라 칼리브레이션을 하는 데에 있어서 더욱 정확한 데이터를 추출할 수 있다. 또한 초기 세팅이 더욱 정확하게 이루어져야 할 것이다. 연계방안으로는 마커가 없는 실시간의 인간 모션 시스템을 구현하는 데에 있다.

지금과 같이 4가지 색깔이 전부 요구되어 지지는 않으며 단일의 피부색으로만 검출하여 사람의 신체를 3차원 모델링 하는 것이다. 예를 들어 사람의 손, 목, 얼굴과 같은 특징점 들을 추출하도록 구현할 것이다.

### 참고 문헌

- [1] Multiple View Geometry in Computer Vision (Second Edition), Richard Hatley, 2003 Second Edition, Chapter 1~4, 6, 7, 9, 11
- [2] A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology using Off-the-shelf TVCameras and Lenses, Roger Y. Tsai, Page 326~344
- [3] Accurate Calibration of Stereo Cameras for Machine Vision, Liangfu Li, Zuren Feng, Yuanjing Feng
- [4] 3 차원 비전 조강현, 유범재, 1998, Chapter 2~7
- [5] Computer Vision and Image Processing Anthony Barrett Computer Vision, Forsyth & Ponce
- [6] Efficient Stereo with Multiple Windowing - Andrea Fusiello, Vito Roberto
- [7] 디지털 영상처리 강동중, 하중은, 2003
- [8] <http://www.ar-vision.com/tech/tech-senser.asp>
- [9] <http://www.git.co.kr/products-1/arial.htm>
- [10] 김효선, 김명희 “적외선, 비컨과 삼각측량법을 이용한 카메라 트래킹”, HCI2002, 2002.2