

# 모션 시뮬레이션을 위한 동영상에서의 3D 위치 정보 추출

박혜선<sup>o</sup> 강신국 박민호\* 김항준  
경북대학교 컴퓨터공학과 인공지능연구실  
경북대학교 전산정보원\*  
(hspark<sup>o</sup>, skkang, kimhj)@ailab.knu.ac.kr  
(mhpark)@knu.ac.kr\*

## 3D Position Information Extraction of Video Image for Motion Simulation

Hye-Sun Park<sup>o</sup> Sin-Kuk Kang Min-Ho Park\* Hang-Joon Kim  
Dept. of Computer Engineering, Kyungpook National University  
Computer Center, Kyungpook National University\*

### 요 약

패턴을 기반으로 한 AR(Augmented Reality) 시스템은 실시간 동영상 안에 가상 물체들을 정확하게 올리기(registering) 위한 좋은 방법이다. AR 시스템을 구현하기 위해서는 우선 카메라가 보고 있는 영상의 3D 위치 정보를 추출하여야 한다. 본 논문에서는 카메라가 보고 있는 체크판 영상의 3D 위치 정보를 자동적으로 추출하여 그것과 동기적으로 움직이는 가상의 object를 구현하는 시스템을 제안한다.

제안된 방법은 카메라 1 대를 가지고 어떠한 sensor나 marker를 사용하지 않고 시간적 정보만을 이용하여 비교적 정확한 3D 위치 정보를 추출할 수 있고, 추출된 3D 위치 정보를 통해 자연스러운 3D 모션 시뮬레이션을 구현할 수 있다.

## 1. 서 론

AR은 실세계 환경과 문자 및 그래픽 등의 부가 정보-가상현실을 실시간으로 합성해 사용자에게 보여주는 기술로써, 최근, 많은 사람들의 관심의 대상이 되고 있는 연구 분야이다. 이러한 기술을 통해 사용자들은 실세계에 대한 이해를 '증강'시키고 '확장'시킬 수 있다. 이러한 AR system은 요즘 게임 분야, 의학 분야 등으로 그 응용이 늘어나고 있다. 미래에는 이와 같은 AR의 기술이 좀더 실생활에 접목하게 될 것이라고 생각된다. 따라서 미래에는 AR의 기술과 wearable computer-몸에 걸칠 수 있는 컴퓨터[1][2]가 결합하여 언제 어디서나 필요한 정보를 사용자의 눈앞에 실시간으로 제공하는 시대가 별로 멀지 않을 것이라고 생각된다.

이러한 AR 시스템을 구현하기 위해서는 우선 카메라가 보고 있는 영상의 3D 위치 정보를 추출하여야 한다. 일반적으로 3D 위치정보를 찾는 방법에서는 Gyro와 같은 sensor를 이용하여 찾거나, 카메라를 통해 영상을 보고 camera calibration을 이용하여 찾는 방법을 주로 한다.

본 논문에서는 고정된 카메라 1 대를 가지고, 어떠한 sensor나 marker를 사용하지 않고 시각적 정보[3]만을 이용하여 카메라가 보고 있는 실제 체크판의 3D 위치 정

보를 찾고, 찾은 정보를 통해 이와 동기적으로 움직이는 가상의 object를 생성하는 시스템을 제안한다.

제안된 시스템은 2D 좌표값으로부터 3D 좌표값을 구하는 방법으로 planar homographies를 사용한다.

먼저 제안된 시스템 구성도에 대해 간단히 설명하고 이에 대한 구체적인 방법을 서술한다.

## 2. 제안된 시스템 구성도

제안된 시스템은 고정된 1 대의 카메라를 통해 보여지는 동영상으로부터 체크판의 특정한 네모 박스를 검출하고 검출된 네모박스의 2D 정보를 기반으로 3D 정보를 추정하여 동영상 안에서, 체크판의 움직임에 따라 동기적으로 동시에 같은 움직임을 가지는 가상의 object를 구현한다.

실제 비디오 영상 내에서 카메라가 object의 패턴(체크판의 가만 사각형)을 보고 그 object(체크판)를 검출하여, 그것의 3D 위치 정보를 추출하고 추출된 정보를 통하여 그 object와 동기적으로 움직이는 가상의 object(3D 사각형이 올려진 가상의 체크판)를 시뮬레이션한다.

제안된 시스템의 구성은 그림 1과 같다.

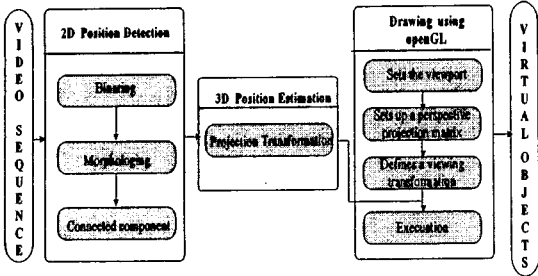


그림 1. 시스템 구성도

### 3. 체스판 검출

우선 입력되는 동영상 안에서 체스판 만을 검출하기 위해, 동영상의 이진화 영상을 구한다. 구한 이진화 영상으로부터 Morphology 연산을 이용하여 연결된 모든 영역들을 따로 떼어낸다. 각각 떼어낸 영역들 안에서 연결 성분 조사(connected component) 방법을 사용하여 체스판 위의 까만 네모 점들만을 검출하여 그 결과, 비디오 입력 영상으로부터 체스판 위의 까만 네모 영역의 2D 중심 좌표가 출력된다.

그림 2 은 실제 사용된 체스판의 영상과 이진 영상, 그리고 그림 3 는 체스판 영상의 2D 좌표값을 추출한 결과를 나타낸 것이다.

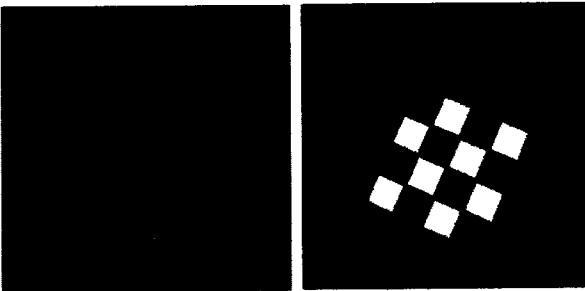


그림 2. 실제 영상과 이진 영상

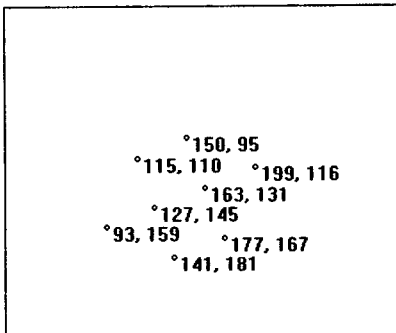


그림 3. 2D 좌표값 추출

### 4. 3D 위치 정보 추출

검출된 체스판의 2D 좌표값으로부터 planar homographies 를 사용하여 3D 위치 정보를 추정한다. 이를 위해, 체스판은 가상의 object 가 놓여질 실세계 좌표로 정의되고 그림 4 와 같이 표현된다.

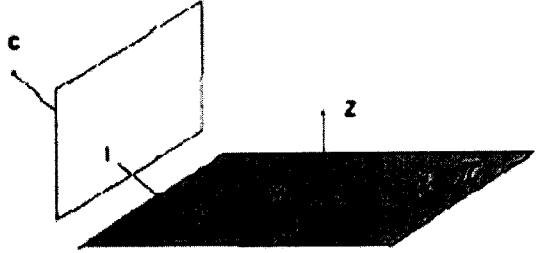


그림 4. 좌표계

$\pi$  로 표현된 체스판(XY-plane)은  $Z=0$  에 위치한 실세계 좌표계라고 가정한 것이다. C(카메라 좌표계)를 통해 보는 실세계에  $I_{\pi}$  좌표값은 I(이미지좌표계)로 매핑된다. 그림 1 은 식 1 과 같이 나타낼 수 있다.

$$I = PI_{\pi} = [P_1 \ P_2 \ P_3 \ P_4] \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = [P_1 \ P_2 \ P_4] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (1)$$

이 때, P 는 3\*4 의 perspective projection matrix 이다. 식 1 을 식 2 와 같이 나타낼 수 있는데, H 는 homograph 로 잘 알려진, 3\*3 의 2D 에서 2D 로 projective transform 이다 [4].

$$[P_1 \ P_2 \ P_4] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (2)$$

제한된 시스템에서는 이 H 를 구하므로써 3D 위치정보를 추출한다. 체스판 위의 점(실세계 좌표값)은 이미지 공간 안에 점과 식 3 에서 주어진 2 개의 linear equations 에 따라 일치된다.

$$\begin{aligned} X_i(h_{31}x_w + h_{32}y_w + h_{33}) &= h_{11}x_w + h_{12}y_w + h_{13} \\ Y_i(h_{31}x_w + h_{32}y_w + h_{33}) &= h_{21}x_w + h_{22}y_w + h_{23} \end{aligned} \quad (3)$$

식 3 의  $H_{ij}$  는 H 의 ij 번째 원소로 표현되고 matrix form 은 다음과 같다.

$$\begin{pmatrix} X_w Y_w 1 & 0 & 0 & 0 & -X_w X_w - X_w Y_w - X_w \\ 0 & 0 & 0 & X_w Y_w 1 & -Y_w X_w - Y_w Y_w - Y_w \end{pmatrix} H = 0 \quad (4)$$

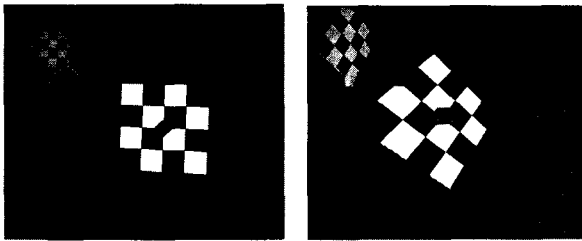
H 는 다음과 같은  $h = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]$ , 9

개 vector 원소로 이루어져 있다. 따라서 적어도  $4 \leq n$  범위 안에서, 패턴 즉 실제계 좌표와 이미지 공간상의 좌표와 일치하는 non-collinear 점을 통해, singular value decomposition 를 사용하여 h의 모든 원소를 구한다 [5].

### 5. 실험 및 결과

제안된 시스템은 추정된 3D 위치 정보를 open GL 를 사용하여 가상의 object 를 그려준다.

실험에 사용된 영상은 실제 체스판을 움직이므로써 발생하는 잡음 및 조도 변화, 영상내의 잡음 등을 포함하고 있다. 실험된 영상은 고정된 카메라 1 대에서 초당 4 프레임으로 획득된 320\*240 크기의 실제 체스판의 영상들이다. 실험은 팬텀III 450Ghz Pc 에서 Visual c++으로 하였다. 한 프레임 당 평균 처리 시간은 0.25 초이다.



(a) 정면 그림

(b) Rotation 한 그림 1

(c) Rotation 한 그림 2

(d) Translation 한 그림

그림 5. 실험 결과

그림 5 는 실제 사용된 체스판의 동영상을 보고 실제 체스판에 따른 가상판을 openGL 로 구현하여 나타낸 영상들이다.

### 6. 결론

본 논문에서는 비디오 영상 내에서 존재하는 object 의 3D 위치 정보를 추출하여 여기에 이와 동기적으로 움직이는 가상 object 를 추가하는 시스템을 구현함으로써 최근 컴퓨터 분야에서 최고의 관심의 대상이 되고 있는 AR(Augmented Reality)를 실현시킬 수 있는 좋은 바탕이 된다. AR 에 대한 연구는 앞으로 실생활에 더욱 근접하게 다가올 것이며, 이러한 점에서 제안된 시스템은 여러 분야에서 응용될 것이다. 따라서, 향후에는 제안된 시스템

으로부터 제안된 가상의 체스판을 기반으로 가상의 공을 올리고 또 그것을 움직이게 한다면 증강된 체스판을 실감(AR 의 실현)할 수 있을 것이다.

본 논문은 카메라를 통한 시각 정보를 활용함으로써 인간이 실생활의 물체를 보고 그 물체를 검출 인식하는 방법 그대로, 카메라가 보고 있는 물체를 어떠한 sensor 도 사용하지 않고 시각 정보만으로 검출하고 검출된 대상의 3 차원 위치 정보를 추정하여 찾아낼 수 있었다.

### 6. 참고 문헌

- [1] Brian Clarkson, "Recognizing User's Context from Wearable Sensors: Baseline System", Vismod, March 4, 2000
- [2] J.Rekimoto and K.Nagao. Agent augmented reality: a software agent meets the world. In proceeding of Second International Conference on Multiagent Systems(ICMAS-96).
- [3] David W. Murray, "Steering and Navigation Behaviours using Fixation", British Medicine Vision Conference'99
- [4] A.Zisserman, R.Hartley. *Multiple view Geometry*. Cambridge University Press, 2000.
- [5] E.Trucco, A.Verri. *Introductory Techniques for 3D Computer Vision*. Prentice-Hall, 1998.