

영상 증강현실 구현을 위한 카메라 자세 추정 방법

Camera Pose Estimation Method for Implementing Vision Augmented Reality

임 현, 이 영 삼
Hyon Lim, Young Sam Lee

402-751 인천광역시 남구 용현동 253 인하대학교 전자전기공학부
alex@alexlab.net, lys@inha.ac.kr

Abstract

The main purpose of Augmented Reality (AR) is to enhance understanding by superimposing additional information on reality. Additional information should be aligned with reality to make immersive environment. To realize this feature of AR, system should estimate user's direction such as eye sight, auditory organ. In particular, visual AR system should estimate user sight direction to superimpose 3D object on reality. To estimate user's sight, camera calibration and pose estimation technique is needed. In this paper, we introduce pose estimation and camera calibration method based on monocular vision. Also experimental results based on this method will be provided.

초 록

증강현실은 실재하는 정보에 부가적인 정보를 정합하여 이해를 증진시키고자 하는데 목적이 있다. 부가적인 정보의 정합은 몰입 환경을 구축하기 위하여 현실과 일치된 시점에서 이루어져야 한다. 이러한 특징을 실현하기 위하여, 시스템은 사람의 시각이나 청각에 대한 지향각을 추정해야 한다. 특히, 영상 증강현실의 경우에는 한 개의 카메라를 기반으로 사용자가 보는 시점을 추정하는 일이 선행되어야만 3차원 정보를 지정된 위치에 정합하여 증강현실의 기능을 구현하는 일이 가능하다. 카메라 자세를 추정하기 위해서는 카메라 교정과 특징점 기반의 카메라 자세추정 기법이 구현되어야 한다. 본 논문에서는 한 개의 카메라를 통하여 촬영 시점의 정보를 알아내는 방법을 소개하고 이에 대한 구현 방법을 소개한다.

Key Words

Augmented Reality, Camera Calibration, Camera Pose Estimation

1. 서론

증강현실은 현실 정보에 부가적인 정보를 정합하여 그 이해를 증진시키고자 하는데 목적이 있다. 이러한 증강현실은 주로 멀티미디어 객체를 현실 정보에 정합하는 형태로 소개되어 왔으며 관련 연구가 활발히 진행되어 왔다. 증강현실은 크게 2가지의 멀티미디어 정보를 부가하는 형태로 연구되어 왔다. 첫째는 영상정보으로써 3차원 컴퓨터 비전 이론을 기반으로 하여 마커(Marker)에 지정된 3차원 객체를 도시하는 것을 목표로 하고 있다. 이러한 영상 증강현실은 ARToolKit [8]이라는 공개 소프트웨어를 통하여 구현되어 왔다. ARToolKit은 카메라 교정, 카메라 자세 추정, OpenGL을 기반으로 3차원 객체를 표현할 수 있는 표현부를 포함하고 있다. 둘째는 음향정보으로써 3차원 공간에 지향성을 가진 음원을 재생하는 것으로 부가적인 정보를 제공하는 것을 목표로 한다. 음향 증강현실은 아직 지원되는 공개 소프트웨어가 없다. 따라서, 영상 증강현실에 비하여 그 연구가 활발하지 않은 편이다. 음향 증강현실은 HRTF(Head Related Transfer Function)을 기반으로 하여 사용자에게 가상의 3차원 음향을 제공한다.

증강현실은 기술이나 이론적이기 보다는 개념적인 것으로 이것을 구현하기 위해서는 그에 해당하는 다른 방법의 도입이 필요하다. 필요한 컴퓨터 비전 이론을 도입하기 위하여 영상 증강현실의 목표를 정리하면 (i) 사용자가 보는 시점을 계산하여 (ii) 3차원 객체가 지정된 위치에 도시하는 것이다. 첫 번째 목표에 대해서는 카메라 교정과 특징점 기반의 카메라 자세추정이 필요하다. 카메라 교정은 처음 한번만 수행되어 고정된 값이며 카메라의 자세는 사용자가 다른 시점을 취할 때마다 바뀌는 값이다. 카메라의 자세를 추정 후에는 이용하는 3차원 객체 표현 라이브러리에 맞도록 그 좌표를 변환해주는 일이 필요하다.

본 논문에서는 영상 증강현실을 구현하기 위하여 필요한 3차원 컴퓨터 비전 이론을 소개한다. 먼저 카메라

교정 방법을 소개하고 이를 기반으로 카메라의 자세를 추정하는 방법을 소개한다. 또한 이러한 카메라 자세 추정 방법을 기반으로 임베디드 장치 기반의 영상 증강현실을 구현한 사례를 소개한다.

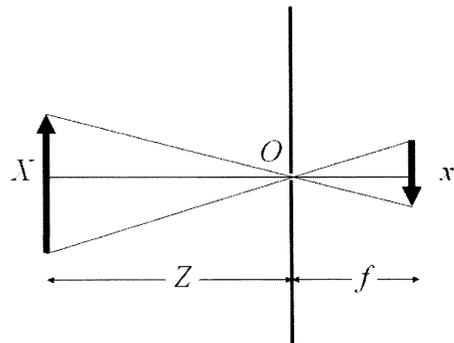


(그림 1) 구현된 임베디드 장치 기반의 영상 증강현실

2. 카메라 교정(Calibration)

2.1 핀홀 카메라 모델

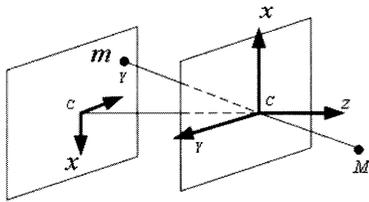
카메라는 빛을 모아 촬상소자(CCD)에 전달하고 이를 디지털 신호로 변환하는 장치로 정의할 수 있다. 카메라는 하나의 변환 함수로써 모델링할 수 있는데 이는 초점거리와 렌즈 왜곡 등이 반영되어야 한다. 먼저, 카메라 모델에 대한 설명을 도입하기 전에 핀홀 카메라(Pinhole Camera) 모델을 통하여 카메라 기하학에 대하여 설명하고자 한다.



(그림 2) 핀홀 카메라 모델

핀홀 카메라는 영상을 작은 구멍을 통하여 관찰하는 것을 모델링 한 것으로 그림 2 과 같다. 이때는 원래 영상과 도립되어 나타나며 그 크기는 다음과 같다.

$$x = -f \frac{X}{Z} \quad (1)$$



(그림 3) 2차원에서의 확장

$$x = -f \frac{X}{Z} \quad y = -f \frac{Y}{Z} \quad (2)$$

2.2 사영 카메라 모델

2차원 영상으로부터 3차원 객체의 위치와 오일러 회전각을 알아내기 위해서는 공간상의 3차원 객체가 2차원 평면인 CCD에 투영되는 관계를 모델링 하여야 한다.

이러한 모델링 결과는 카메라 보정(calibration) 작업에 이용된다. 카메라 보정은 렌즈 왜곡, 초점거리, CCD의 한 점(pixel)의 가로/세로 비율을 구하는 것을 말한다. 카메라 보정 단계에서 얻어지는 변수는 내부변수(intrinsic parameter)로 후에 카메라의 자세를 추정하는데 고정 변수의 역할을 담당한다.

비선형 식 (2)는 다음과 같은 선형 표현으로 나타낼 수 있다.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (3)$$

식 (3)에서 (X_c, Y_c, Z_c) 는 카메라 좌표계 상의 좌표이며 (u, v) 는 CCD에 투영된 2차원 영상좌표이다. 주어진 벡터 $m = [u, v]^T$ 에 대하여 확장벡터 $\tilde{m} = [u, v, 1]^T$ 를 정의하면 식 (3)은 다음과 같이 바꿀 수 있다.

$$s \tilde{m} = P \tilde{M} \quad (4)$$

식 (4)에서 로 표현되는 식은 식 (3)에서의 사영 행렬과 같다

$$P = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

2차원 좌표인 영상 좌표계로부터 실세계 좌표를 얻기 위해서는 카메라 좌표계를 먼저 회전 시키고, 자세를 유지한 채로 평행이동 시킨다. 회전은 3X3의 행렬 R 이며 평행 이동인 병진 행렬은 로 t 표현한다. 이 두 변수 R, t 를 카메라의 외부 변수라고 하며 본 논문에서 추정하고자 하는 값이다.

어떤 점에서의 카메라 좌표계에 의한 좌표를 M_c 라고 하고 세계 좌표계의 좌표를 M_w 라고 하자. 이들의 관계는 다음과 같이 표현 가능하다.

$$\tilde{M}_c = R \tilde{M}_w + t \quad (5)$$

또는 $\tilde{M}_c = D \tilde{M}_w$ 로 나타낸다. 이 식에서 D 는 유클리드 좌표 상에서의 강체 변환(rigidbody transformation) 행렬로 다음과 같다.

$$D = \begin{bmatrix} R_{3 \times 3} & t \\ O_3^t & 1 \end{bmatrix} \quad O_3 = [0, 0, 0]^T \quad (6)$$

식 (3)과 식 (5)로부터 식 (7)을 얻을 수 있으며 새로운 사영행렬을 식 (8)과 같이 정의 할 수 있다.

$$s \tilde{m} = P \tilde{M}_c = P D \tilde{M}_w \quad (7)$$

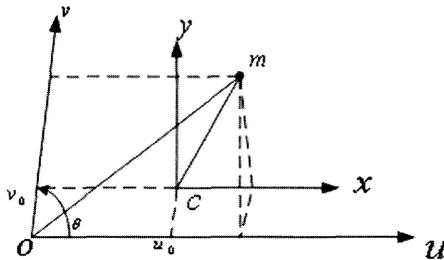
$$P_w = PD = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} \\ O_3^T & 1 \end{bmatrix} \quad (8)$$

2.3 카메라의 기구적 왜곡의 고려

실제 카메라가 공장에서 조립되어 출시 될 때에는 여러 가지 불확실한 요소로 인하여 이상적인 카메라 모델과는 일치하지 않는 경우가 있다. 따라서 이러한 파라미터를 고려하지 않으면 오차가 발생한다.

카메라가 문제점을 가질 수 있는 부분을 요약하면 다음과 같다.

- (1) 영상의 광축(optical axis) 중심이 CCD의 중심과 일치하지 않는다.
- (2) 영상의 두 좌표축의 비율이 다르다.
- (3) 영상의 두 좌표축이 직교하지 않는다.



(그림 4) 카메라 내부 변수의 오차 관계

좌표계 o-uv와 좌표계 c-xy의 좌표를 각각 $m = [u, v]^T$ 와 $m_s = [x, y]^T$ 로 정의 하면 다음과 같은 관계식이 성립한다.

$$\tilde{m} = H\tilde{m}_s \quad (9)$$

$$H = \begin{bmatrix} k_u - k_u \cot \theta & u_0 \\ 0 & k_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

식 (10)에서의 변수는 실제 카메라의 수평축과 수직축이 이루는 각도는 θ , 각각의 축의 단위길이는 k_u, k_v 영상의 중심(Optical axis)는 u_0, v_0 이다. 식 (4)에 따라 $s\tilde{m}_s = P\tilde{M}$ 이므로 카메라의 왜곡이 고려된 사영행렬 P_{new} 는 식 (2)와 같다.

$$P_{new} = HP = \begin{bmatrix} fk_u - fk_u \cot \theta & u_0 & 0 \\ 0 & fk_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

식(11)에서 fk_u 와 fk_v 는 항상 곱의 형태로 나타나므로 이를 같이 모델링하여 α_u 와 α_v 로 정의한다. 카메라의 내부변수는 $\alpha_u, \alpha_v, \theta, u_0, v_0$ 의 5가지이다. 식 (11)을 단위행렬로 분리해내면 다음과 같다.

$$\begin{bmatrix} \alpha_u & \alpha_u \cot \theta & u_0 \\ 0 & \alpha_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = K [L_3 | O_3] \quad (12)$$

식 (12)에서 변수 K를 카메라 내부 변수(Intrinsic parameter)라고 하며 5자유도(DOF)를 가진다. 또한 K는 상삼각 행렬이다.

카메라의 외부 변수(Extrinsic parameter)와 내부변수를 결합한 사영행렬은 다음과 같이 정의 할 수 있다.

$$P = \begin{bmatrix} \alpha_u & \alpha_u \cot \theta & u_0 \\ 0 & \alpha_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} \\ O_3^T & 1 \end{bmatrix} \quad (13)$$

실세계 좌표(World Coordinate) \tilde{M}_w 가 영상좌표 \tilde{m} 로 변환되는 과정은 다음과 같다.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (14)$$

2.4 선형 카메라 교정(Calibration) 법

카메라 교정(calibration)은 3차원 비전 처리를 위하여 카메라의 내부변수를 구해내는 과정을 말한다. 하지

만 카메라 교정의 이론을 활용하여 외부변수도 추정 가능하므로 이에 대한 이론은 중요하다고 할 수 있다. 카메라를 교정하는 방법은 크게 다음의 두가지로 분류할 수 있다.

- (1) 이미 알고 있는 패턴을 이용한 카메라 교정 방법 [2,3]
- (2) 영상간의 상호성 (correlation)을 이용한 자체 교정 방법 [4,5,6]

이미 알고있는 패턴을 이용한 카메라 교정방법은 치수가 미리 알려진 인식이 수월한 물체를 활용하여 그 형태를 관찰함으로써 카메라의 내부 변수를 추정해내는 과정을 말한다. 이 방법은 정확도가 높기 때문에 일반적으로 사용되고 있다. 처음 등장한 형태는 2개 이상의 강판을 서로 수직이 되도록 구성하여 좌표축을 구성하고 이를 촬영한 영상에서 교차점을 추출해내어 카메라 교정을 하였지만 근래에는 단순히 1장의 강판을 이용하여 여러장의 영상을 기반으로 카메라 교정 문제를 해결하고 있다. 이를 위하여 Faugeras[1], Tsai[2], Zhang[3]의 방법이 많이 사용되고 있다.

교정패턴을 활용하지 않는 것은 자체교정 (self-calibration)으로 알려져 있으며 Luong 과 Maybank 등의 연구가 있다.

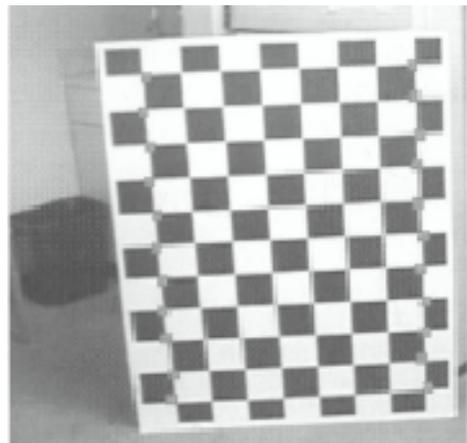
이 방법은 임의의 동영상으로부터 각 프레임에 대한 대응점을 구해내고 이에 대한 관계를 이용하여 내부 변수를 추정한다. 하지만 이 방법은 특징점을 추출하는 방법에 의존적이고 그 정확도가 교정패턴을 활용하는 방법보다 낮은 단점이 있다. 근래에는 Scale Invariant Feature Transform[6]을 이용하여 신뢰도 있는 두 영상간의 대응점 일치법으로 자체교정을 수행하기도 한다.

$$p = [m_{11}, m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{31}, m_{32}, m_{33}, m_{34}]^T$$

$$L = \begin{bmatrix} X_W^1 & Y_W^1 & Z_W^1 & 1 & 0 & 0 & 0 & -u_1 X_W^1 - u_1 Y_W^1 - u_1 Z_W^1 - u_1 \\ 0 & 0 & 0 & 0 & X_W^1 & Y_W^1 & Z_W^1 & 1 - v_1 X_W^1 - v_1 Y_W^1 - v_1 Z_W^1 - v_1 \\ \vdots & \vdots \\ X_W^n & Y_W^n & Z_W^n & 1 & 0 & 0 & 0 & -u_n X_W^n - u_n Y_W^n - u_n Z_W^n - u_n \\ 0 & 0 & 0 & 0 & X_W^n & Y_W^n & Z_W^n & 1 - v_n X_W^n - v_n Y_W^n - v_n Z_W^n - v_n \end{bmatrix}$$



(그림 5) 2개 이상의 판을 수직배치 하는 방법



(그림 6) Tsai, Zhang의 방법

카메라 교정은 크게 2단계로 나눌 수 있다. 선형 해법을 이용하여 초기치를 구한 후 비선형 해법을 적용하여 신뢰성 있는 교정 계수를 구해내는 방법이다. 선형 해법은 반복적인 계산수행을 하지 않아 계산량이 적은 반면 그 정확도는 낮다. 따라서, 선형 해법은 이론에 대한 이해를 돕거나 비선형 해법을 위한 초기치를 산정할 때 사용한다. 식(14)로부터 3차원 세계좌표 \bar{M}_W 와 2차원 영상좌표 \tilde{m} 간의 관계를 이용하여 다음의 선형 방정식을 세울 수 있다.

$$\begin{aligned} m_1^T P_{world}^i - u_i m_3^T P_{world}^i + m_{14} - u_i m_{34} &= 0 \\ m_2^T P_{world}^i - u_i m_3^T P_{world}^i + m_{24} - v_i m_{34} &= 0 \end{aligned}$$

$$M = K [L_3 | O_3] \begin{bmatrix} R_{3 \times 3} & t \\ O_3^T & 1 \end{bmatrix} = \begin{bmatrix} m_1^T & m_{14} \\ m_2^T & m_{24} \\ m_3^T & m_{34} \end{bmatrix} \quad (15)$$

이로부터 다음의 선형동차방정식 (Linear Homogeneous Equation) 을 세울 수 있다. L 은 의 세계좌표 집합 행렬이며, p 는 세계좌표를 2차원 평면영상으로 투영하는 투영행렬이다.

$$Lp = 0 \quad (16)$$

식 (16)은 특이치 분해 (Singular Value Decomposition) 을 이용하여 구할 수 있다.

$$L = U \Sigma V^T = \sum_{i=1}^{12} \sigma_i u_i v_i^T \quad (17)$$

이때의 p 는 가장 작은 σ 값에 대응하는 v 이다. 이 때, 다음의 두가지 경우가 가능하다.

- (1) 구하려는 해의 개수 (n=12) 와 점이 정확히 일치 할 경우. 이 경우는 6점만 대응 하면된다.
- (2) 구하려는 해의 개수보다 점이 더 많을 경우.

이 경우는 $L^T L$ 의 최소 고유값에 대응하는 벡터가 p 의 최근접 해가 된다. 이 해는 p 의 비선형 해법의 초기치로 가장 적당하다.

3. 카메라 자세 추정

3.1 평면체 기반의 자세 추정

카메라의 자세를 추정하기 위하여 적절한 평면체를 영상으로부터 추적해 내는 일이 선형 되어야 한다. 평

면체를 추적하지 않고도 카메라의 자세를 추정하는 일은 가능하다. 하지만, 다량의 영상에서 대응점을 구해 내는 일은 불확실의 요소를 포함하고 있다. 따라서, 미지의 대응점 보다영상에서 보다 추적하기 쉬운 형태의 평면을 찾아내는 일이 좀더 강건한 자세추정을 가능하게 한다. 많은 연구에서 이러한 평면체를 마커 (Marker) 라고 부르고 있으며 그 형태는 다양하다. 영상처리의 부하를 줄이기 위하여 많은 종류의 마커들은 흰 바탕에 검은 형태를 띄고 있다.



(그림 7) 마커의 한 형태 (HITLabNZ)

Simon[7]은 연속된 영상간의 Homography¹⁾ 를 계산함으로써 카메라 자세추정 문제를 해결하였다. 2절에서 설명한 카메라 교정을 통하여 각 영상에서의 Homography 를 추정해 낸 후 이로부터 카메라 자세를 얻는다. Simon 의 알고리즘은 그림 7과 같다.

평면체의 경우 세계좌표를 평면체의 일정 점으로 정의하면 $Z=0$ 의 관계를 얻을 수 있다. 따라서 세계좌표 $\bar{M}_w = [X, Y, 0, 1]^T$ 를 얻을 수 있다. \bar{M}_w 가 \tilde{m} 로 투영되는 관계를 재정의하면 다음과 같은 식을 얻을 수 있다.

*Homography 는 일종의 변환행렬로 간주할 수 있다. 즉, 한 좌표계에서 정의된 점을 다른 한 좌표계로 이동할 수 있도록 정의된 행렬이다.

초기화:

1. 초기 영상에서 평면체를 찾는다.
2. 평면체에서 특징점을 추출한다.
3. 카메라 행렬 K 를 초기화 한다.

이후, H 를 $frame_i$ 와 $frame_{i+1}$ 로부터 추정한다:

1. N 개의 특징점을 $frame_{i+1}$ 로 부터 추출한다. 이는 $\{x_j^{i+1}\}_{j=1}^N$ 을 정의해준다.
2. $frame_i$ 로 부터 $i+1$ 사이의 대응점을 정의한다. 이는 $x_j^i \leftrightarrow x_k^{i+1}$ 의 대응 관계를 설정한다.
3. 2번 과정으로 부터 계산된 대응관계로 부터 Homography H_i^{i+1} 을 구한다.
4. H_i^{i+1} 로 부터 카메라 자세를 추정한다.

(그림 8 Simon 의 알고리즘)

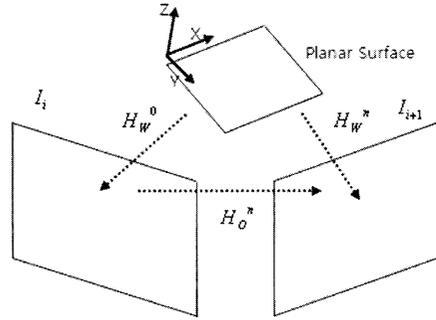
$$P_{\infty} = p^i \widetilde{M}_W = K [r_1^i \ r_2^i \ r_3^i \ t^i] \begin{bmatrix} X_W \\ Y_W \\ 0 \\ 1 \end{bmatrix}$$

$$= K [r_1^i \ r_2^i \ t^i] \begin{bmatrix} X_W \\ Y_W \\ 1 \end{bmatrix} = H_W^i \begin{bmatrix} X_W \\ Y_W \\ 1 \end{bmatrix} \quad (17)$$

식 (17)에서 H_W^i 는 i 번째 이미지의 특징점 \widetilde{M}_W 를 2차원 영상에서의 관찰점 \widetilde{m} 로 변환해주는 변환행렬로 간주 할 수 있다. 관찰된 영상에서 다시 세계좌표로 변환하는 것은 $H_W^{-1} = H^W$ 로 가능하다. 하지만 세계좌표를 직접 정의하는것은 불가능하다. 따라서 두 영상간의 Homography 를 다음의 식으로 정의한다.

$$H_i^{i+1} = H_W^{i+1} (H_W^i)^{-1} \quad (18)$$

이로부터 직접 세계좌표를 정의하지 않고도 두 영상간의 대응점을 통하여 H_i^{i+1} 을 구해내면 각각의 영상에 대한 Homography 를 구할 수 있다.



(그림 9) 두 이미지 간의 Homography 개념도

3.2 영상간의 Homography 추정

두 영상간의 Homography는 i 번째 영상에서의 특징점 집합 $x^i = (x, y, 1)$ 와 $i+1$ 번째 영상에서의 특징점 집합 $x^{i+1} = (x', y', 1)$ 로 부터 구할 수 있다. 이는 2.4절에서 기술한 선형 방정식과 동일한 형태로 해결 할 수 있다. 그 식은 다음과 같다.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \lambda H_i^{i+1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (19)$$

식 (19)로 부터 다음의 선형 방정식을 세울 수 있다.

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

이를 통하여 선형동차방정식 $Ah=0$ 을 얻어 낼 수 있다. h 는 행렬 A의 Null space 이므로 Singular Value Decomposition 을 통하여 계산 할 수 있다. 행렬 A 는 다음과 같다.

$$A = \begin{bmatrix} x_1 y_1 1 0 0 0 -x'_1 x_1 -x'_1 y_1 -x'_1 \\ 0 0 0 x_1 y_1 1 -y'_1 x_1 -y'_1 y_1 -y'_1 \\ x_2 y_2 1 0 0 0 -x'_2 x_2 -x'_2 y_2 -x'_2 \\ 0 0 0 x_2 y_2 1 -y'_2 x_2 -y'_2 y_2 -y'_2 \\ x_3 y_3 1 0 0 0 -x'_3 x_3 -x'_3 y_3 -x'_3 \\ 0 0 0 x_3 y_3 1 -y'_3 x_3 -y'_3 y_3 -y'_3 \\ x_4 y_4 1 0 0 0 -x'_4 x_4 -x'_4 y_4 -x'_4 \\ 0 0 0 x_4 y_4 1 -y'_4 x_4 -y'_4 y_4 -y'_4 \end{bmatrix}$$

3.3 외부 파라미터 추정

3.2 장에서 소개한 방법을 통하여 두 영상간의 대응점을 통한 Homography 를 계산한 후 외부 파라미터를 추정 할 수 있다. 외부 파라미터는 3차원 객체를 현실 영상에 정합하기 위한 중요한 변수이다.

세계좌표로부터 현재 이미지 좌표로의 Homography 는 $H_{ii}^i = H_0^i H_{ii}^0$ 로 정의 할 수 있다. 하지만, 이를 직접적으로 구하는 것은 앞서 말한 직접 세계좌표를 정의해야 하는 문제가 있다. 따라서, 다량의 영상간의 대응점을 기반으로 이를 재구성 할 수 있다. 그 내용은 다음과 같다.

$$H_0^i = H_{i-1}^i H_{i-2}^{i-1} \dots H_0^1$$

따라서, 다량의 영상간의 대응점을 계속 구해 나간다면, H_{ii}^i 도 구할 수 있다.

$H_{ii}^i = K[r_1^i \ r_2^i \ t_i]$ 의 관계로부터 회전 행렬 요소 r_1, r_2 를 구할 수 있다. 먼저, 행렬 K 는 2절에서 소개한 카메라 교정 방법을 통하여 구해놓는다. 회전요소 r_1, r_2 는 $K^{-1}H_{ii}^i$ 의 처음 두 열로부터 구할 수 있다. 각각의 회전요소는 Orthonormal 하기때문에 r_3 은 r_1, r_2 의 외적으로 구할 수 있다. 따라서, 완벽히 추정된 p^i 는 다음과 같다.

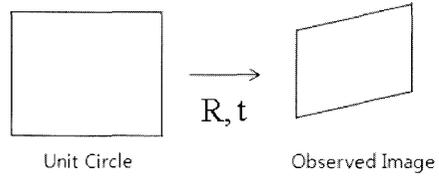
$$p^i = K[r_1^i \ r_2^i \ (r_1^i \times r_2^i) \ t^i] \quad (20)$$

3.3 H_{ii}^0 구하기

3차원 객체를 정확히 정렬하기 위하여 H_{ii}^i 를 구하는 것은 필수적이다. 이를 위하여 H_0^i 와 H_{ii}^0 를 구해야 하는데 3.3 절에서 H_0^i 를 구하는 방법은 이미 소개하였다. 따라서 본 절에서는 H_{ii}^0 를 구하는 방법을 소개한다.

H_{ii}^0 는 전체 알고리즘의 초기치에 해당하는데, 이것은 카메라 교정에서의 외부 파라미터 추정과 유사하다.

H_{ii}^0 는 Metric과 Projective 요소로 나눌 수 있다. Metric 요소는 회전, 이동, 스케일 변환을 함축하고 있다.



(그림 9) 단위원의 Metric 투영에 의한 변형

H 를 추정하기 위하여 단위 사각형 (0,0),(1,0),(1,1),(0,1)을 정의하고 이를 각각 (x_k^W, y_k^W) 로 나타낸다. 그다음 (x_k^W, y_k^W) 를 추출된 평면의 4-점 (x_k', y_k') 으로 변환하는 행렬 H 를 추정한다. 둘의 관계는 다음과 같다. H 를 구하는 방법은 3.2 장을 참조한다.

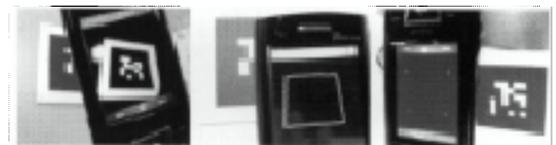
$$x_k' = \lambda_k H x_k^W$$

$$H = H_{ii}^0 D = K[r_1 \ r_2 \ t] \begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (21)$$

$$K^{-1}H = [r_1 \ sr_2 \ t]$$

K 는 카메라 교정을 통하여 이미 구해졌으므로 식 (21) 을 통하여 r_1, r_2 및 s 를 구할 수 있다. 이를 구하는 방법은 다음과 같다.

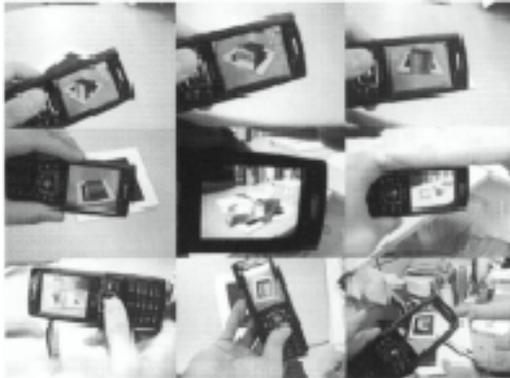
1. 단위 사각형을 (x_k', y_k') 으로 대응하는 H 를 구한다.
2. $s = \left\| \frac{h_2}{h_1} \right\|$ 를 이용하여 s 를 구한다.
3. $H_{ii}^0 = HD^{-1} = Hdiag(1, \frac{1}{s}, 1)$ 를 구한다.



(그림 10) 단위사각형을 대응하기 위한 코너점 추출 실험 영상

3.4 외부 파라미터 추정

2절과 3.3절에 걸친 내용을 스마트폰에 구현하여 실험하였다. 구현된 스마트폰은 400Mhz 의 성능을 가진 ARM10 프로세서 이다.



(그림 11) 다양한 각도에서 마커를 추적하여 3차원 객체를 도시하는 실험 영상

영상 배열		
A	B	C
D	E	F
G	H	I

그림 10은 OpenGL/ES와 3절에서 설명한 알고리즘을 기반으로 제작한 소프트웨어의 실험 영상이다. 다양한 각도에서 촬영하여도 객체가 지정된 각도로 도시됨을 알 수 있다. 실험은 총 2가지 모델에 대하여 이루어졌다. 정방형의 상자 모양의 모델은 A-B-C-D-G-H-I의 순서로 촬영 되었으며 두개의 직사각형 상자의 교차모델은 E-F 순서로 촬영 되었다

3. 결론 및 향후 연구방향

본 논문에서는 카메라 교정법과 평면체 자세추정 방법을 이용하여 영상 증강현실을 구현해 보았다. 카메라 교정법은 영상증강현실 뿐만 아니라 촬영된 영상을 기반으로 3차원 정보를 얻어내고자 할 때 중요한 역할을

담당한다. 또한, 선형해법을 통한 대응점 기반의 카메라 자세추정 방법은 영상 기반의 3차원 재구성에도 활용 될 수 있다.

본 논문을 통하여 구현한 3차원 자세추정 방법은 아직 마커를 기반으로 하고 있다. 마커를 이용한 방법은 상대적으로 안정적인 장점이 있으나 사용자로 하여금 몰입형 환경을 구축하는데 방해요인이 되어 왔다. 따라서 우리는 향후 연구방향으로 SIFT [6]와 같은 특징점 추출기를 이용하여 영상의 자체교정을 통하여 마커가 없이도 동작하는 3차원 카메라 자세추정기의 제작을 진행하고자 한다.

Acknowledgement

이 논문은 2007년도 정부(과학기술부)의 재원으로 한국 과학재단의 지원을 받아 수행된 연구임 (No. R01-2007-000-10651-0)

참고문헌

- [1] O. Faugeras, Three-Dimensional Computer Vision: a Geometric Viewpoint. MIT Press, 1993.
- [2] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses), IEEE Journal of Robotics and Automation, 3(4): 323-344, 1987.
- [3] Z.Zhang, A flexible new technique for camera calibration, Technical Report MSR-TR-98-71, Microsoft Corporation, 1998.
- [4] Q.T. Luong and O. Faugeras, Self-calibration of a moving camera from point correspondences and fundamental matrices, International Journal of Computer Vision, 22(3):261-289, 1997.[5] 여인춘 외 “임베디드 프로그래밍” 길벗, 2004
- [5] S.J. Maybank and O.Faugeras, A theory of self-calibration of a moving camera, International Journal of Computer Vision, 8(2): 123-152, 1992.
- [6] David G. Lowe, Distinctive image features from scale-

invariant keypoints, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[7] Simon G, Fitzgibbon A, and Zisserman A, Markerless Tracking using Planar Structures in the Scene, Proc. International Symposium on Augmented Reality, 2000.

[8] ARToolKit - HITLabNZ
<http://www.hitl.washington.edu/artoolkit>

■ 저자 소개

◆ 임 현



• 2007년 현재 인하대학교 전자전기공학부 학사과정 재학
<주요 관심 분야: 3차원 컴퓨터 비전, 임베디드 시스템, 로봇 머신비전, 컴퓨터 제어>

◆ 이영삼



• 1997년 2월 인하대학교 전기공학과 (학사)
• 1999년 2월 인하대학교 전기공학과 (석사)
• 2003년 8월 서울대학교 전기컴퓨터공학부 (박사)
• 2004년 9월 ~ 현재 : 인하대학교 전자전기공학부 조교수
<주요 관심 분야: 임베디드 시스템, 디지털 제어, CACSD (Computer-aided Control System Design) 소프트웨어>