

다중 영상으로부터 3차원 재구성

김상훈*, 김태은**, 최중수*

*중앙대학교 첨단영상대학원 영상공학과

**남서울대학교 공학부 멀티미디어학과

3D Reconstruction from multiple Images

Sang-Hoon Kim*, Tae-Eun Kim**, Jong-Soo Choi*

*Dept. of Image Engineering,

Graduate School of Advanced Imaging Science, Multimedia and Film,,

Chung-Ang Univ.

**Dept. of Multimedia, NamSeoul Univ.

요 약

본 논문에서는 3차원 재구성에 있어서 필수 불가결한 기술인 카메라 교정 방법에 있어서 특정 교정 물체나 또는 영상에서의 제약 조건 등을 요구하지 않고 영상 내에 산재되어 있는 기하학 정보를 이용하여 카메라 내부 파라미터를 추출하고 영상간의 카메라 움직임울 계산하여 3차원 재구성하는 알고리즘을 제안한다. 공간에서의 직교하는 평행선들의 집합이 만들어 낸 각 축 방향으로의 3개의 소실 점을 이용하면 그 투영 영상에 대한 카메라 내부 파라미터를 얻을 수 있게 된다. 또 한 영상간의 대응점 관계를 이용하면 두 영상 사이의 상대적인 카메라의 회전 및 이동 성분을 얻어 낼 수 있다. 따라서 카메라의 내부, 외부 성분을 추출함으로써 사영 행렬을 계산하고 역 투영 방법에 의해서 3차원 재구성을 구현하게 된다.

I 서론

최근 데스크탑 컴퓨터 또는 웹 브라우저 상에서 3차원 그래픽이 광목하게 증가함으로써 값비싼 CAD 와 같은 장비의 사용 없이 현실감 있는 3차원 모델의 수요가 날로 증가하고 있다. 따라서, 현실 세계의 사진과 같은 현실감 있는 3차원 모델을 만들기 위한 연구가 많이 진행되어왔다. 그러나 합성 영상에 대한 꾸준한 진전에도 불구하고 복잡한 물체에 대한 3차원 모델링은 힘들고 오랜 시간을 요구하는 작업이다. 가장 일반적인 모델링 프로그램은 Maya, SoftImage, 3D Studio등이 있는데 영상의 현실감은 질적인 면과 3D 모델의 복잡성에 따라서 감소하게 된다. 모델링에 있어서 최근의 경향은 실제 물체에 대한 사진으로부터 3차원을 재구성하는 것과 실제 정확한 정보를 이용하고자 하는 것이다. 이러한 새로운 패러다임은 영상기반 모델링(IBM: Image Base Modeling)이라고 불리는데 일반적으로 실제 사진 영상을 이용한 모델링 기법이다.[1] 가장 큰 장점은 실제 세계로부터 취득한 텍스처를 이용하여 좀더 실제에 가까운 영상을 만들어 내는 것이다. 영상으로부터 3차원 정보를 얻기 위해서는 카메라 교정이라는 과정을 필수적으로 사용하는데, 이러한 카메라 교정 기술은 로보틱스(Robotics), 3차원 비전(3D Vision)등의 다른 여러 분야에도 많이 활용되고

있는 기술로 그에 따른 상당한 연구가 진행되어 왔다. 그 중 카메라 교정 장비 및 3차원 위치 센서 등을 이용하는 방법이 있으나 이는 교정을 위한 고가의 장비 및 제한된 취득 환경을 요구하는 단점이 있다. 카메라 이외의 다른 장비를 사용하지 않고 취득한 영상만을 이용하는 시각 기반(vision-based) 방법 중에는 Z. Zhang은 사진에 알고 있는 체크 패턴을 실세계에 포함하여 그 패턴이 투영된 영상을 이용하여 카메라를 교정하였다.[2] 이 방법은 비교적 정확한 카메라 파라미터를 얻어 낼 수는 있으나 영상 내에 항상 사진에 알고 있는 패턴이 존재해야 하는 전제 조건이 존재한다. 따라서, 본 논문에서는 소실 점을 이용해 단순하면서도 직관적인 해법을 통하여 내부 파라미터를 추출하고 두 장의 영상으로부터 상대적인 움직임 추정을 통해서 3차원 재구성을 구현하였다.

II 기하학 정보를 이용한 내부 카메라 교정

공간상에 존재하는 평행한 선분들은 카메라를 통해 2차원 영상에 투영될 때 하나의 소실 점을 형성하게 된다. 이러한 소실 점들은 매우 강력한 기하학적 정보이다. 이것들은 선들의 방향성과 평면들의 방위에 대한 많은 정보들을 전달한다. 이러한 소실 점은 카메라와 장면 시점과의 명백

한 기하학적 정보 없이 영상으로부터 직접적으로 계산할 수 있다. 때로 이러한 소실 점들은 영상 밖에 놓여질 수 있으나 계산에는 아무런 영향을 미치지 않는다.[3][5][6]

2.1 소실 점의 계산(Computing vanishing points)

실세계에서의 평행한 선들은 각각의 소실 점에서 교차한다. 그러므로, 소실 점은 적어도 그러한 2개 이상의 선들로 정의되어진다. 그러나 2개 이상의 선들을 이용할 수 있는 경우, 본 논문에서 소실 점을 찾기 위해 Maximum Likelihood Estimate(MLE) 알고리즘을 적용 하였다. 따라서 소실 점을 계산하기 위해서 실세계에서 서로 평행한 선들의 벡터 곱을 통해서 계산할 수 있다. 사영공간에서 두 직선

$$l_1 = [a_1, b_1, c_1]^T \quad (1)$$

$$l_2 = [a_2, b_2, c_2]^T \quad (2)$$

의 교점은 $l_1 \wedge l_2 = l_1 \times l_2$ 이다.

2.1 선 조정.

소실 점을 구하기 위해 초기 사용자에 의해서 평행한 선들의 집합을 정의해 주어야 한다. 이 때 발생할 수 있는 오차를 최소화하기 위해 선 조정 알고리즘을 적용하였다. 카메라의 투영된 선 영상은 일반적으로 많은 잡음과 선에 속하지 않는 점(outlier)을 포함하기 때문에 기존의 최소 자승법(Least squares method)으로는 정확한 선의 방정식을 얻기 힘들다. 최소 자승법은 자승 거리의 합을 최소화하는 파라미터를 찾는 것으로 몇 개의 점이 큰 거리를 가지고 있을 때 큰 오류를 범할 수 있다. 이를 극복하기 위해서 본 논문은 가중치 된 최소 자승(weighted least squares)을 이용한 최대 유사 추정법(maximum likelihood estimation method)을 적용하였다.

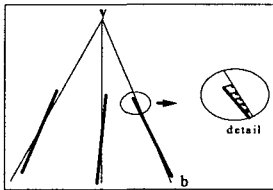


그림 1. 선 조정(Line Fitting)

선 조정 알고리즘

1. Least square 방법을 이용하여 초기 직선의 방정식을 구한다.

$$y = a_1x + a_0 \quad (3)$$

2. 최소 자승 오차의 합을 최소화하는 파라미터를 구한다.

$$E_k(a_0, a_1) = \sum_{i=1}^n [y_i - (a_1x_i + a_0)]^2 \quad (4)$$

3. 각 점에서부터 직선까지의 거리를 구한다.

4. 거리에 따라 가중치를 구한다.

5. 가중치 된 최소 자승법을 이용하여 직선을 구한다.

$$E_k(a_0, a_1) = \sum_{i=1}^n W(d_i^{k+1}) [y_i - (a_1x_i + a_0)]^2 \quad (6)$$

위의 식은 반복횟수 k와 지난 과정에서의 최소치 합과 가중치 함수 W에 의해 최소화되는 파라미터를 반복적으로 수행하여 찾는 방법이다. 따라서 구하고자 하는 파라미터의 선형 함수에 의한 행렬의 가장 작은 고유 값(eigenvalue)에 해당하는 고유 벡터(eigenvector)가 E_k 를 최소화하는 해가 된다. d는 선과 점의 수직 거리이며 δ 는 Welsch의 $\delta=2.9846$ 으로 설정하였다.[4]

2.2 카메라 내부 파라미터 추정.

서로 직교하는 3방향성에 대한 무한 점을 고려하면, 각 x, y, z축 방향의 무한 점이 사영 행렬에 의해 소실 점으로 이미지에 투영되는 관계식을 유도할 수 있다.[3][8]

$$\begin{bmatrix} \lambda_1 u_1 & \lambda_2 u_2 & \lambda_3 u_3 \\ \lambda_1 v_1 & \lambda_2 v_2 & \lambda_3 v_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix} = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (7)$$

여기서, u_i, v_i 는 소실 점을 그리고, λ_i 는 초기 스케일(scale)성분을 나타낸다. 위의 식은 카메라 교정 행렬(Camera Calibration Matrix : K)과 회전 행렬(Rotation Matrix : R)로 다시 표현 될 수 있다.

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = KR \quad (8)$$

카메라 교정 행렬에서 카메라 영상에서 각 x, y축에 대한 크기 비율을 알고 두 축의 기울기(s)값을 영으로 가정하면 식(9)은 다음과 같이 나타낼 수 있다.

$$R = \begin{bmatrix} \lambda_1(u_1 - u_0)/\alpha & \lambda_2(u_2 - u_0)/\alpha & \lambda_3(u_3 - u_0)/\alpha \\ \lambda_1(v_1 - v_0)/\alpha & \lambda_2(v_2 - v_0)/\alpha & \lambda_3(v_3 - v_0)/\alpha \\ \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix} \quad (9)$$

$x_i=(u_i, v_i)$ 로 정의하고 회전 행렬의 직교 성을 이용하여 카메라 중심과 카메라 투영 평면(image plane)이 직교하는 중심점 좌표(principal point) u_0 와 v_0 , 그리고 스케일 성분 λ_i 를 구할 수 있다.

$$(x_1 - x_0) \cdot (x_2 - x_3) = 0 \quad (10)$$

$$(X_2 - X_0) \cdot (X_1 - X_3) = 0 \quad (11)$$

$$(X_3 - X_0) \cdot (X_2 - X_1) = 0 \quad (12)$$

식 (10),(11),(12)를 이용하여 X_0 에 대하여 $AX=0$ 의 선형 방정식으로 표현할 수 있다. X_0 의 동차 좌표계에 대하여 SVD를 이용하면 X_0 (principal point)의 해를 구할 수 있다. 그리고 식(13)은 α^2 을 구하는데 이용된다.[5][6] 따라서 영상에서의 초점거리(focal length)는 다음과 같다.

$$\alpha = \sqrt{(X_0 - X_1) \cdot (X_3 - X_0)} \quad (13)$$

III 대응점 관계를 이용한 외부 카메라 교정

이 장에서는 대응되는 두 장의 영상으로부터 대응점을 이용한 외부 카메라 성분을 찾는 방법에 대해 논한다.

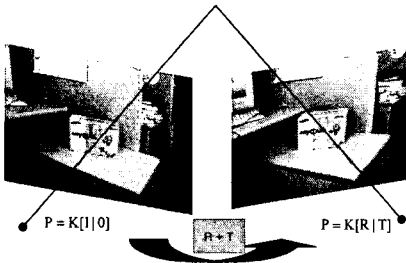


그림 2. 영상간의 상대적인 움직임

3.1 외부 카메라 교정

외부 카메라 교정을 위해서는 필수 행렬을 이용한다. 필수 행렬은 기본 행렬로부터 유도될 수 있는데, 필수 행렬(essential matrix)은 기본 행렬(fundamental matrix)의 특별한 형태로 일반화된 영상 좌표의 경우에 해당된다. 카메라 내부 파라미터를 알고 있을 경우 카메라 행렬의 역행렬을 이용하여 영상 좌표 x 에 적용하면

$$\hat{x} = K^{-1}x \quad (14)$$

일반화된 좌표로 표현되는 영상 좌표를 얻을 수 있다. 필수 행렬(essential matrix)의 정의로부터 다음과 같은 수식으로 표현 가능하다.

$$\hat{x}^T E \hat{x} = 0 \quad (15)$$

이는 영상 사이의 대응되는 점 $x \leftrightarrow x'$ 에 대한 일반화된 영상 좌표의 형태로 표현한 것이다. 식 (14)를 식 (15)에 대입하면

$$x^T K^{-T} E K^{-1} x = 0 \quad (16)$$

을 얻을 수 있는데 $E = K^T F K$ 의 관계를 적용하면 필수 행렬과 기본 행렬의 관계를 정의 할 수 있다. 식 (16)로부터 내부 카메라 성분을 알고 대응점으로부터 기본 행렬을 구하면 필수 행렬을 계산할 수 있다.[7] 이렇게 구한 필수 행렬을 이용하면 두 영상간의 회전과 이동 성분을 추출할 수 있다. 필수 행렬의 특성은 같은 2개의 유일치(singular value)와 하나의 영의 값을 갖는다. 이는 두 개의 행렬의 곱

으로 표현 가능한데 첫째는 비대칭 행렬이고 다른 하나는 직교 행렬이 되겠다. 유일치(singular value)에 대한 분해(decomposition) 결과는 $(\lambda, \lambda, 0)$ 이다. 행렬 A 와 B를 다음과 같이 정의하면 필수 행렬(essential matrix)에 대한 SVD는

$$E = USV^T = \lambda \cdot UABV^T \quad (17)$$

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$E = \lambda \cdot UAU^T \cdot UB^T \quad \text{or} \quad E = -\lambda \cdot UAU^T \cdot UB^T \quad (18)$$

식 (18)의 마지막 직교 부분이 회전 성분이 된다. 따라서 4개의 가능한 해가 존재하는데 이중 2개는 우리가 바라는 행렬식(determinant)이 1이 아닌 -1의 값을 가지므로 제거될 수 있다.

$$R = UB^T \quad \text{or} \quad R = -UB^T \quad (19)$$

$$R = UB^T V^T \quad \text{or} \quad R = -UB^T V^T \quad (20)$$

또한 필수 행렬은 다음의 식을 만족하는데 이는 정확한 이동 성분에 대한 해답을 제시해 준다.

$$E^T \cdot (t_x \ t_y \ t_z)^T = 0 \quad (21)$$

$$t = (t_x \ t_y \ 1)^T \quad \text{or} \quad t = (-t_x \ -t_y \ -1)^T \quad (22)$$

3.4 역 투영법에 의한 3차원 좌표 추출

카메라에 대한 내부, 외부 파라미터가 계산되어지면 역 투영법에 의해서 2차원 영상에 대한 3차원 정보를 구할 수 있다. 각 영상에 대해서 $x=PX$, $x'=P'X$ 관계식은 $AX=0$ 의 선형 방정식으로 표현 가능한데 과정은 다음과 같다. 각 영상의 점에서 다음의 세 개의 식을 유도하기 위해 먼저 스케일 성분을 외적의 연산을 통해서 제거 할 수 있다. 첫 영상에 대해서 $x \times (PX) = 0$ 으로부터

$$x(P^{3T}X) - (P^{1T}X) = 0 \quad (23)$$

$$y(P^{3T}X) - (P^{2T}X) = 0 \quad (24)$$

$$x(P^{2T}X) - y(P^{1T}X) = 0 \quad (25)$$

여기서 P^{iT} 는 사영 행렬 P의 행 벡터이다. 이 식들은 X성분에 대해서 다음의 선형 방정식으로 분해가 가능하다.[7]

$$AX = 0 \quad (26)$$

$$A = \begin{bmatrix} xP^{3T} - P^{1T} \\ yP^{3T} - P^{2T} \\ x'P^{3T} - P^{1T} \\ y'P^{3T} - P^{2T} \end{bmatrix} \quad (27)$$

식 (27)은 DLT(Direct Linear transformation) 알고리즘을 사용해서 동차 방법(Homogeneous Method)으로 구할 수 있는데, A 행렬의 SVD(Singular value decomposition)을 사용해서 최소의 유일치를 갖는 유일 벡터가 곧 방정식의 해가 되고 우리가 구하고자 하는 공간상의 좌표가 되겠다.

IV 실험 및 결과

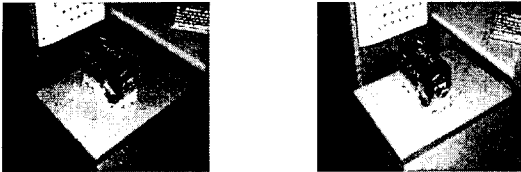


그림 3. 원 영상 (640*512)

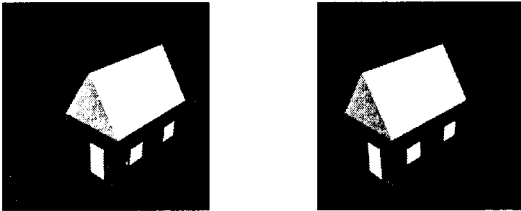


그림 4. 원 영상 (500*500)

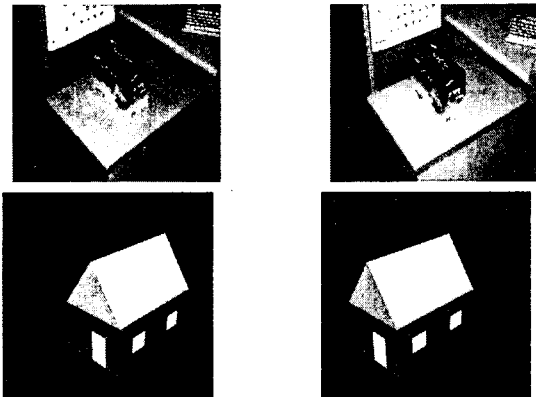


그림 5. 선 조정 알고리즘을 적용한 영상

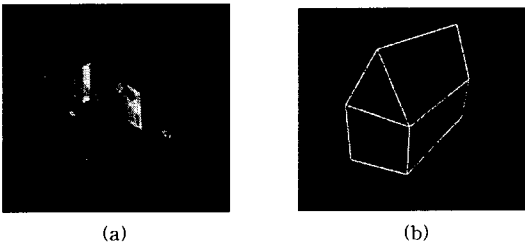


그림 6. OpenGL로 구현한 3차원 모델 영상

그림. 3 4 는 카메라 시점을 달리 한 다중 영상이다.

그림. 5는 선 조정 알고리즘을 적용하기 위해 에지(edge)를 찾고 이 에지를 따라서 가장치 된 최소 자승(weighted least squares)을 이용한 최대 유사 추정법(maximum likelihood estimation method)을 적용하여 새로운 선들로 조정하였다. 그림 6.은 3차원 공간상에 Texture를 입혀서 모델링한 결과이다.

V 결론 및 향후 과제

본 논문에서는 영상내의 기하학 정보를 이용하여 카메라 내부 성분을 추출하고 영상간의 정확한 대응점 정합 알고리즘을 이용하여 카메라 교정을 하였다. 이를 기반으로 2차원 영상으로부터 3차원 정보를 추출하였고 3차원 재구성을 실현하였다. 초기 영상 입력 과정에서 사용자에게 의한 평행선들을 지정함으로써 발생할 수 있는 오차를 줄이기 위해 선 조정 과정을 도입하여, 단순 하지만 직관적인 기하학 정보를 이용하는데 있어서 오차를 최대한 줄이도록 하였다. 또한 카메라 움직임 추정 부분에 있어서는 영상간의 명확한 대응점 관계를 위해서 RANSAC 방법을 도입하여 보다 정확한 카메라 움직임 정보를 추출하였다. 결국 일반적인 사영 카메라 모델에서 카메라 내부, 외부 정보를 얻을 수 있었으며 이를 통해서 3차원 정보를 얻어 낼 수 있었다. 그러나 본 시스템에서는 카메라 내부 파라미터를 구하기 위해 초기 사용자 입력이 들어가야 하는 제약이 따른다. 그러므로 향후 과제로 입력 영상으로부터 자동적으로 소실 점을 검출할 수 있도록 하는 연구가 진행되어야 하겠다.

REFERENCES

- [1] P. Debevec, C. Taylor, and J. Malik. "Modeling and rendering architecture from photos: a hybrid geometry and image-base approach", SIGGRAPH'96 Conference proceedings, Annual Conference Series, pp.11-20, July,1996.
- [2] Zhengyou Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.22, no.11, pp.1-20, 1998.
- [3] Antonio Criminisi, Ian D. Reid, Andrew Zisserman: Single View Metrology. International Journal of Computer Vision vol.40, no.2, pp.123-148, 2000
- [4] Z. Zhang, "Parameter Estimation Techniques : A Tutorial with Application to Conic Fitting", The Journal of Image and Vision Computing, vol.15, no.1, pp.59-76, 1997.
- [5] B. Caprile and V. Torre. "Using Vanishing Points for Camera Calibration", International Journal of Computer Vision, vol.4 pp.127-139, 1990.
- [6]Roberto Cipolla, Duncan Robertson," PhotoBuilder-3D Models of Architectural Scenes from Uncalibrated Images", ICMCS99, Italy,1999.
- [7] R. Hartly and A. Zisserman, "Multiple view geometry in computer vision", Cambridge University Press, pp.184-243, 2000.
- [8] D. Leibowitz, A. Criminisi and A. Zisserman, "Creating architectural models from images" EuroGraphics, Milan, Italy Sep. 1999.