단일 비디오 카메라를 이용한 3차원 구조의 조밀한 복원

박정우[O] 박종승* 황용구 이만재
한국정보통신대학교 디지털미디어연구소, 인천대학교 컴퓨터공학과*
jwpark@icu.ac.kr

3D Dense Surface Reconstruction from Single-Camera Video

Jungwoo Park[O], Jong-Seung Park*, Yong K. Hwang, Manjai Lee
Digital Media Lab, ICU, University of Incheon*

요 약

이 논문은 한 대의 카메라에서 얻은 일련의 영상을 해석하여 단순한 2차원의 영상을 3차원물체로 복원하는 방법에 대해 설명을 한다. 이러한 3차원 복원 방법은 카메라 내부 변수가 동일하다는 가정을 이용하여 별도의 캘리브레이션 작업 없이 한 대의 카메라로부터 얻은 여러 장의 영상을 이용한다. 이 논문에서 제안한 방법은 내부 변수 중 카메라 행렬의 단순화와 사영 기하를 이용한 것이다. 이 방법은 실제 비디오 프레임에 가상의 그래픽 모델을 더하는 AR (Augmented reality) 분야에 특히 유용하다. 이 논문에서의 실험은 실제 여러 비디오 스트림 데이터를 바탕으로 수행되었고, 하나의 카메라를 사용한 동영상에서 3차원 구조로 복원하는 실험 결과는 시스템의 유용성을 보여준다.

1. 서 론

최근 한대의 카메라에서 얻은 영상을 사용하여 3차원 영상으로 재구성하는 몇몇의 기초적인 연구가 진행되어 왔다. Kahl 등[1] 은 임의의 휴대용 비디오카메라의 움직임을 모델링하는 방법을 제시하였고, MAP(maximum a posteriori)를 응용하여 유클리디안 기하에서의 구조와 모션을 측정하는 방법을 제안하였다.

이 논문에서는 비디오카메라 한 대를 가지고 3차원 공간 구조를 복원하는 방법을 소개한다. 현재 기술의 특징은 특정 조건하에서는 정확도가 높은 최적의 해를 제공한다는 것이다. 그래서 우리가 이용한 조건은 가장 일반적인 것으로 모든 영상들은 동일한 조건을 가진, 즉, 초점이나 줌의 변화가 없는 카메라에서 얻은 것이라 가정한다. 즉, 현재 존재하는 이론이나 알고리즘은 카메라 내부 파라미터가 변하지 않는 조건에서 얻어진 것이라 가정한다. 영상의 주축은 직교한다고 가정하고, 종횡비도 이미 알고 있다고 가정한다.

두 장의 영상에서 신뢰 높은 픽셀의 상관관계를 구하는 것은 어려운 일이다. 특히 이것이 비디오 시퀀스와 같은 여러 장의 영상일 때는 더욱 그렇다. 또한 특징을 추적하는 기법은 짧은 순간에 갑자기 커다랗게 변하는 모션이나, 영상의 교합 그리고 영상내의 존재하는 다수개의 애매한 특징점들 때문에 종종 실패한다. 또한 이 기법은 한 프레임에서 발생한 에러가 그 다음 프레임들로 전파되는 높은 확률을 가진다. 그러나 이러한 문제점들은 아웃라이어 제거 기법을 통해 개선이 가능하다. 그래서 이 방법을 사용해서 특징점을 추적하면 많은 에러를 감소시킬 수 있다. 카메라 파라미터나 에피폴라 기하의 지식을 사용하여 이러한 문제들을 해결할 수 있다.

매트릭 캘리브레이션을 계산하기 전에 아핀 캘리브레이션의 계산이 선행되어야 한다. 이와 동일하게, 먼저 사영 공간에서의 무한 평면을 계산하여야 하는데, 이것이 가장 어려운 과정 중의 하나이다. Armstrong 등[2]은 단지 시점의 변화만을 이용해서 아핀 캘리브레인션을 얻을 수 있었다. 거기에 일반적인 모션의 시점을 더하여 매트릭 캘리브레이션은 쉽게 얻을 수 있다. Hartley[3]는 dense search를 이용한 알고리즘을 제안했고, Pollefeys와 Gool [4]은 4차식의 사영 공간의 무한 평면 방정식의 계수를 이용한 비선형 알고리즘을 제시했다.

우리가 제한한 3차원 복원 시스템은 전처리 과정, 여러 시점 분석과 3차원 복원의 3가지 방법으로 나뉜다. 그림 1은 우리가 제한한 시스템의 순서를 다이어그램으로 나타낸 것이다.

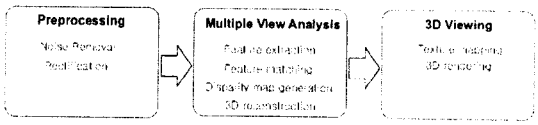


그림 1. 3차원 구조 복원 시스템 다이어그램

2. 카메라 파라미터와 사영 매트릭스

카메라가 움직이면 영상의 카메라 좌표가 변하기 때문에 이러한 카메라의 움직임을 계산하기 위해서는 카메라 회전변환 행렬 R과 이동 벡터 t를 고려해야 한다 [3]. 그림 2에서는 카메라 움직임에 따른 외부 파라미터의 상관관계를 보여준다.

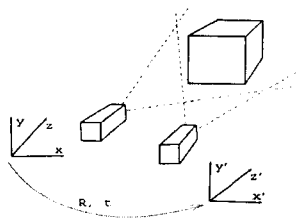


그림 2. 카메라 이동과 외부 파라미터

영상의 포인트 M'이 다른 위치의 포인트 M으로 변환된다고 가정할 때, 한 영상의 포인트 모션은 다음과 같이 모델링 할 수 있다.

$$\mathbf{M}=\begin{bmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0}_3^T & 1\end{bmatrix}\mathbf{M}'$$

여기서 R은 회전 행렬, $\mathbf{t}=[t_x\ t_y\ t_z]^T$은 이동 벡터이다. 카메라 모션은 영상 모션과 역의 관계가 성립하므로, 위의 식을 다음의 등가의 식으로 표현할 수 있다.

$$\mathbf{M}'=\begin{bmatrix}\mathbf{R}^T & -\mathbf{R}^T\mathbf{t}\\ \mathbf{0}_3^T & 1\end{bmatrix}\mathbf{M}$$

우리가 제안한 카메라 모델이 실제 카메라 영상의 중심과 사영된 영상의 중심이 같고, 사영된 영상이 XY평면과 수평인 Z = 1인 축 위에 존재한다면, 동차 좌표 포인트 $\mathbf{M}=[X\ Y\ Z\ 1]^T$와 이에 대응하는 동차 이미지 포인트 $\mathbf{m}_R=[x\ y\ 1]^T$에 대해서 프로젝션 과정은 다음과 같이 모델링 될 수 있다.

$$\mathbf{m}_R\sim[I_{3\times3}|\mathbf{0}_3]^T$$

실제 응용 프로그램에서 카메라의 초점 거리 f는 1이 아니므로, 3차원에서 2차원으로의 프로젝션은 f만큼 고려해야 한다.

$$\mathbf{m}\sim\mathrm{K}\mathbf{m}_R$$

여기서,

$$\mathrm{K}\sim\begin{bmatrix}f_x & 0 & c_x\\ 0 & f_y & c_y\\ 0 & 0 & 1\end{bmatrix} \quad (1)$$

인 관계가 성립하고, f_x와 f_y는 CCD의 한 픽셀의 가로와 높이가 고려된 초점 거리이다. 그리고 (c_x, c_y)는 이미지 센서 배열의 중앙을 나타낸다.
실제 3차원 세계가 영상 평면으로 프로젝션 되는 것으로 카메라 내부 파라미터 캘리브레이션과 카메라의 위치와 회전된 정도를 계산할 수 있다. 이러한 프로젝션은 다음과 같은 식으로 나타낼 수 있다.

$$\mathbf{m}\sim\mathbf{P}\mathbf{M} \quad (2)$$

여기서 P는 3차원 월드 좌표를 2차원 이미지 좌표로 투영하는 3×4 카메라 프로젝션 행렬을 의미한다. 이것은 동차 좌표계에서 3차원 월드 포인트 $\mathbf{M}=[X\ Y\ Z\ 1]^T$ 가 카메라에 의해 2차원 이미지의 한 포인트 $\mathbf{m}=[x\ y\ 1]^T$ 로 투영되는 관계를 설명하는 것이다. 또한 '~'의 의미는 0을 양변에 곱하는 것을 제외하고 양변이 등가라는 것을 나타낸다. 그래서 전체 프로젝션 행렬은 다음과 같이 카메라 캘리브레이션 행렬, 투영 행렬 그리고 카메라 모션 행렬로 구성된다.

$$\mathbf{m}\sim\mathrm{K}[\mathbf{R}^T\ -\mathbf{R}^T\mathbf{t}]\mathbf{M} \quad (3)$$

식 (3)에서 파라미터 K와 R과 t에 따라 전체 프로젝션되는 과정이 결정된다.

3. 특징점 매칭

3차원 모델을 복원하기 위한 비디오 영상 분석 과정은 힘든 작업 중의 하나이다. 이러한 주요 원인으로는 분석 과정의 초기에 사용할 수 있는 정보가 거의 없다는 것이다. 가장 기본적인 가정으로 핀홀 카메라 모델을 기준으로 하여 영상에 존재하는 물체는 난반사 한다는 것을 제외하고 월드 좌표계나 카메라 좌표계조차 알 수 없다.

이미지 시퀀스를 이용하여 3차원 구조를 복원하기 위하여, 두 개의 연속적인 영상 사이의 특징점을 매칭 하는 문제부터 해결해야 한다. 그러나 만약 영상 사이에 특별한 조건만 만족한다면 두 영상간의 대응되는 포인트나 매칭 가능한 특징점들을 자동으로 찾아내는 것은 간단히 해결될 수 있다. 이를 위한 유용한 가정들로는 두 장의 영상은 동일한 조명하에서 거의 비슷하고 짧은 시간 동안 물체의 움직임은 거의 발생하지 않는다는 것이다. 즉, 짧은 시간동안 물체의 움직임이 크지 않기 때문에 거의 비슷한 포즈가 유지되므로 특징점들은 이전 프레임과 비슷한 위치에 존재하게 된다. 그래서 이러한 경우, 특징점의 위치와 특징점 주위의 명암의 분포가 두 장의 영상에서 비슷하게 나타날 것이다. 또한, 매칭할 수 있는 점이나 특징점들이 그것 주변 픽셀들의 정보와 확연히 다르다면 그러한 점이나 특징점들을 찾는 시간을 크게 줄여줄 수 있고, 영암의 크로스 코릴레이션을 사용하여 특징점들을 찾을 수 있다. 이러한 크로스 코릴레이션은 얼마나 두 패턴이 밀접한 관련이 있는지 측정하는 표준 알고리즘이다. 이 논문에서는 대응되는 두 영상의 특징점들을 찾기 위해 이 기법을 사용하였다. 영상의 한 점 (i,j)에서 정규화되지 않은 상관관계 계수는 다음의 식과 같이 나타낼 수 있다.

$$r(i,j)=\sum_{j=-N_j/2}^{j<N_j/2}\sum_{i=-N_i/2}^{i<N_i/2}(\mathrm{C}_{(i+N_i/2,j+N_j/2)}-\bar{\mathrm{C}})(\mathrm{I}_{(i+i,j+j)}-\bar{\mathrm{I}}) \quad (4)$$

여기서 $\bar{\mathrm{C}}$는 찾고자 하는 특징점이 존재하는 마스크 픽셀들의 평균이고, $\bar{\mathrm{I}}$는 마스크 픽셀과 비교할 다른 영상의 평균을 의미한다.

4. 3차원 복원

비디오카메라에서 얻은 2장 이상의 연속적인 프레임 사이에는 특별한 상관관계가 존재한다. 그것은 오직 특징점 매칭을 통해서만 사영된 3차원 구조를 복원할 수 있다. 이와 같은 특징과 공간 상의 한 포인트와 이에 대응하는 카메라 영상면의 한 포인트 사이의 관계를 구하는 캘리브레이션을 이용하여 2차원의 카메라 영상에서 3차원 물체로의 복원이 가능해진다. 카메라 프로젝션 행렬과 3차원 공간 좌표는 동시에 복원이 된다는 것은 3차원 복원문제라고 알려져 있다. 다음과 같은 관계 때문에

$$\mathbf{P}\mathbf{M}=(\mathbf{P}\mathrm{T}^{-1})(\mathrm{T}\mathbf{M})=\mathbf{P}'\mathbf{M}'$$

복원된 $\mathbf{P}'$와 $\mathbf{M}'$은 임의의 3차원 변환행렬 T에 의해 실제의 P와 공간상의 M은 다를 수 있다. 복원된 $\mathbf{P}'$와 $\mathbf{M}'$은 또한 자연스럽게 다음과 같은 관계가 성립한다.

$$\mathbf{P}'\sim\mathbf{P}\mathrm{T}^{-1},\ \mathbf{M}'\sim\mathrm{T}\mathbf{M}$$

여기서 T는 0이 아닌 4×4의 역행렬이 존재하는 행렬이다.

캘리브레이션 과정에 존재하는 변환은 3개의 종류로 나눌 수 있다. 만약 T가 3차원 공간에서 프로젝션 변환

과정에 속한다면 대응되는 캘리브레이션은 $(P_P^i \mathbf{M}_P^i)$인 프로젝션 복원을 한다. 만약 T가 3차원 공간상의 아핀 변환과정에 포함된다면 대응되는 캘리브레이션은 $(P_A^i, \mathbf{M}_A^i)$인 아핀 복원을 한다. 또한 T가 매트릭 변환과정에 속한다면 대응되는 캘리브레이션은 $(P_M^i \mathbf{M}_M^i)$인 매트릭 복원을 한다. 일단 P_M^i를 알면, 카메라 캘리브레이션 행렬 K^i를 QR 분해과정을 이용해서 직접적으로 얻을 수 있다. 사영 공간에서의 3차원 복원은 카메라 내부 파라미터에 관한 어떠한 제약도 없이 단지 특징점 매칭만을 이용하여 쉽게 얻을 수 있다. 부가적인 제약을 이용해서 사영 캘리브레이션을 아핀 캘리브레이션이나 더 나아가 매트릭 캘리브레이션으로 변환시킬 수 있다. 이러한 성질은 각각의 변환 행렬 T_{PA}과 T_{AM}로 표현 가능하다. T_{PA}는 사영 구조를 아핀 구조를 변화시키고, T_{AM}은 아핀 구조를 매트릭 구조로 변화시킨다. 이러한 3종류의 관계는 다음과 같이 요약된다.

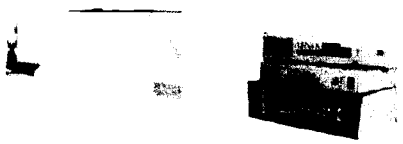
$$P_A^i = P_P^i T_{PA}^{-1},\ P_M^i = P_A^i T_{AM}^{-1} \qquad (5)$$

우리의 궁극적인 목표는 사영 공간에서의 3차원 복원으로부터 매트릭 복원에 대한 적절한 변환 행렬 T_{PA}와 T_{AM}을 찾는 것이다.

복원된 구조를 사영 공간에서 매트릭 공간으로 변화시키려면 8개의 파라미터를 구하는 것만으로 충분하다. p 벡터의 3개의 파라미터와 K의 5개 파라미터가 그것인데, 여기서 p 벡터는 무한 평면에서의 처음 3개의 파라미터가 상수인 $\pi_\infty = [\mathbf{p}^T, 1]^T$이다. 매트릭 공간으로 변환 시킨다는 것은 무한 평면과 앱솔루트 코닉을 정의하는 것과 등가이다. 각각의 카메라 캘리브레이션 K^i와 첫 번째 카메라와 관련된 회전 행렬 R^i과 이동 행렬 $\mathbf{t}^i$ 는 매트릭 공간에서의 3차원 복원을 결정한다. 우리는 사영 변환에서 첫 번째 시선에 대한 정규화 카메라 행렬을 $P^1 = [I|0]$가 되도록 선택했다. 또한 $P_P^i = [A^i|\mathbf{a}^i]$는 사영 복원된 카메라들이라 정의한다. 만약 모든 카메라가 동일한 내부 파라미터들을 가진다는 가정 하에서, $A = K^1$, $\mathbf{t} = \mathbf{0}$라 한다면

$$KK^T = (A^i - \mathbf{a}^i\mathbf{p}^T)KK^T(A^i - \mathbf{a}^i\mathbf{p}^T)^T,\ i = 2, \ldots, m \qquad (6)$$

이 성립함을 알 수 있다.



(a) (b)

그림 3. 그림 10의 영상의 이용하여 3차원으로 재구성한 구조: (a) 조밀한 디스패리티 맵, (b) 텍스쳐를 매핑한 모델

첫 번째 프레임을 제외한 각각의 시점은 5개의 제약을 가지므로 최소 3개 이상의 다른 시점을 가지면 3차원으로 복원된 영상을 얻을 수 있다 [3].

그림 3은 책 영상들로부터 3차원으로 복원한 결과를 보여주는데, (a)는 dense disparity map 이고, (b)는 복원된 3차원 구조에 영상 텍스쳐를 매핑한 결과이다. 그림 4에서는 3차원으로 복원한 쌓여진 책들을 여러 시점에서 본 것이다.

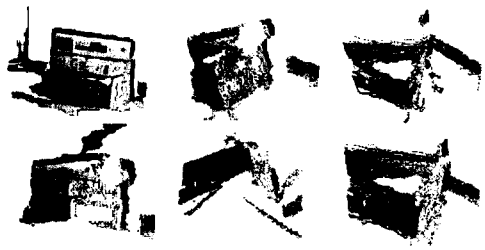


그림 4. 그림 3의 3차원으로 복원한 구조를 여러 방향에서 본 그림

5. 결 론

이 논문에서는 연속적인 비디오카메라 영상에서 3차원으로 복원하는 시스템을 제안하였다. 시스템은 어떠한 카메라 캘리브레이션 과정 없이 단지 카메라 파라미터의 단순화와 사영 기하를 사용하여 구성되었다.

우리가 제한한 방법은 일반 데스크탑 컴퓨터 환경에서 구현 되었고, 이를 바탕으로 실험을 수행하였다. 또한 연속된 영상 프레임에서 3차원 구조를 복원하는 시스템의 성능은 우수하게 측정되었다.

실제로 이러한 시스템은 많은 실용적인 응용 프로그램에 사용될 수 있다. 예를 들면, 건축물의 가상 복원 시스템을 위한 가장 기본이 되는 코어 모듈이나 비디오 영상으로부터 실제 환경을 모델링하는 기계나 크기를 재는 측정기, 그리고 카메라 위치 제어 시스템과 같은 여러 응용 분야에 사용할 수 있다. 특히 제한한 기법은 실사에 특정한 그래픽 모델을 집어넣는 AR 분야에 특히 유용하다.

향후 보완사항으로는 좀 더 정확한 3차원 모델을 위해 복원된 조밀한 표면의 홀을 제거하는 것이다.

참고문헌

1. Kahl, F., Heyden, A.: Euclidean reconstruction and auto-calibration from continuous motion, ICCV 2001, pp. 572-577, 2001
2. Armstrong, M., Zisserman, A., Beardsley, P., Euclidean structure from uncalibrated images, BMVC94, pp. 509-518, 1994
3. Hartley, R., Zisserman, A., Multiple View Geometry in Computer Vision, Cambridge University Press, 2000
4. Pollefeys, M., Gool, L., Stratified self-calibration with the modulus constraint, IEEE T-PAMI 21(8), pp. 707-724, 1999