

분해법기반 프로젝티브 재구성에 관한 연구

정 윤 용, 조 청 운, 홍 현 기
중앙대학교 첨단영상대학원 영상공학과 그래픽스 및 미디어 연구실
전화 : 02-820-5748 / 핸드폰 : 011-9291-3238

A Study on Projective Reconstruction based on Factorization Method

Yoon-Yong Jung, Chung-Woon Jo, and Hyun-ki Hong
Dept. of Image Eng., Graduate School of Advanced Imaging Science Multimedia & Film,
Chung-Ang University
E-Mail : kburngae@hotmail.com, blue@cglab.cse.cau.ac.kr, honghk@cau.ac.kr

Abstract

The recovery of 3D scene structure from multiple views has been long one of the central problems in computer vision. This paper presents a new projective reconstruction method based on factorization for un-calibrated image sequences. The proposed algorithm provides an effective measure to construct frame groups by using various information between frames. The experimental results show that the proposed method can reconstruct a more precise 3D structure than the precious methods such as the merging method.

I. 서론

최근 컴퓨터 비전 분야에서는 여러 장의 영상으로부터 기하학적인 모델링 정보를 해석하는 기술이 활발하게 연구되고 있다. 이러한 3차원 재구성기술은 건축, 가상환경, 의료분야, 영화 등을 포함한 많은 분야에 응용될 수 있다. 더욱이 인터넷의 발달에 따라 매우 현실적이고 실감적인 온라인 대화 시스템이 요구되고 있으며, 3차원 가상환경에 대한 수요역시 매년 급격하게 증가하고 있다. 이러한 환경을 보다 효과적으로 지원하기 위해 영상기반 관련 기술의 개발이 활발히 진행 중이며, 특히 비디오와 같은 비교정(un-calibrated) 영상 시퀀스로부터 카메라의 움직임과 대상 장면의 구조정보를 해석하는 역구가 1990년대 이후로 꾸준히 발표되고 있다.

3차원 재구성과정은 크게 세 단계로 분류된다. 첫째, 카메라의 특성 및 위치정보를 사전 보정이나 패턴 없이 강건하게 구하는 과정, 조밀한 3차원 정보를 충실하게 복원하기 위한 조밀한 매칭(dense matching)과정, 그리고 여러 장의 영상에 대해 보다 정확한 사영행렬을 구하는 과정이다. 본 논문에서는 분해법(factorization)기반의 프로젝티브(projective) 재구성 방법이 영상 시퀀스에 효과적으로 적용되지 못하였던 단점을 보완하기 위해 효과적인 프로젝티브 재구성 알고리즘을 제안하였으며, 실험을 통해 기존의 병합(merging)기반 방법과 그 성능을 비교하였다.

논문의 구성은 다음과 같다. 2장에서 사전정보나 패턴 없이 카메라 특성 및 위치정보를 구하기 위한 카메라의 자동보정(auto-calibration)과 프로젝티브 재구성에 대해 설명하고, 분해법을 이용하여 프로젝티브 재구성을 하기 위해 영상 그룹핑의 기준을 3장에서 제안하였으며, 4장에서 유클리드 공간상의 카메라 특성 및 위치정보를 구하는 과정을 기술하였다. 그리고 위의 과정을 통해 3차원

재구성된 결과를 5장에, 결론 및 앞으로의 연구진행 방향을 6장에 기술하였다.

II. 카메라 자동보정 및 프로젝티브 재구성

1. 카메라의 자동보정

카메라 자동보정 방법은 프로젝티브 재구성 및 유클리드 재구성 과정의 두 단계로 나뉜다. 프로젝티브 재구성이란 구하고자 하는 유클리드 공간상의 3차원 구조정보와 4x4 행렬로 되는 구조정보의 복원이다. 사전정보 없이 영상만으로 대상 장면의 구조를 재구성을 하기 위해 이러한 프로젝티브 재구성 과정이 필요하며, 이 방법은 분해법기반과 병합기반의 방법으로 구분된다[1].

프로젝티브 구조로부터 공간 사영변환행렬을 구하면, 유클리드 구조 복원이 가능하다. 또한 카메라 투영행렬도 복원 할 수 있기 때문에 내부 및 외부 파라미터를 추정할 수 있다. 그러므로 카메라의 자동 보정은 프로젝티브 재구성과 프로젝티브 구조로부터 오차가 적은 공간사영변환행렬을 추출하는 것이 중요하며, 여기서 공간 사영변환행렬은 절대2차원추곡면(absolute quadric)의 추정을 통해 계산된다[2].

2. 프로젝티브 재구성

대상영상에 대한 3차원 구조를 정확하게 재구성하기 위해 오차가 적은 프로젝티브 재구성이 필수적이다. 병합 방법은 첫 번째 카메라를 세계좌표계와 일치시키고 기본행렬(fundamental matrix)을 이용하여 두 번째 카메라의 투영행렬을 추정한다. 첫 번째 카메라와 두 번째 카메라로부터 얻어진 프로젝티브 공간상의 3차원 구조정보를 이용하여 세 번째 카메라 투영행렬을 구한다. 이런 방식으로 비디오와 같은 긴 영상 시퀀스에 대하여 계속 병합하기 때문에 중간프레임에서 에러가 발생하면 그 이후 영상

의 해석에 매우 큰 영향을 미친다. 또한 초기값에 민감하고 에러값의 누적에 쉽게 대응하지 못한다[3].4].

이에 반해 분해법 기반방법은 한번에 카메라 투영행렬과 3차원 구조정보를 계산하기 때문에 병합기반방법에 비해 에러값의 누적에 대응할 수 있고 계산속도도 빠르다[5]. 그러나 영상 해석에 사용되는 일치점이 모든 영상 시퀀스에 공통적으로 존재해야 하는 가장 때문에 비디오 영상 같은 긴 영상에 대하여 적합하지 못하며, 이러한 문제를 해결하기 위해서 병합기반방법을 바탕으로 하는 계층적 프로젝티브 재구성 방법이 제안되었다[1]. 하지만 계산시간이 오래 걸리고 분해법기반 방법에 비해 강건하지 않은 단점이 있다. 본 논문에서는 강건하고 계산시간이 빠른 분해법기반의 프로젝티브 재구성 방법을 제안하였으며 병합기반방법과 그 성능을 비교하였다.

실세계의 점 X 가 카메라투영행렬 P 에 의해 영상의 점 x 에 투영된다면 식(1)과 같이 표기할 수 있다.

$$\lambda_i^j x_i^j = P^j X_j, \quad (1)$$

여기서 i 는 영상의 숫자를 나타내고 j 는 점의 숫자를 나타낸다. n 개의 영상에 대하여 m 개의 일치점이 존재한다면 다음과 같이 하나의 행렬식으로 표기할 수 있다.

$$\begin{pmatrix} \lambda_1^1 x_1^1 & \lambda_2^1 x_2^1 & \dots & \lambda_n^1 x_n^1 \\ \lambda_1^2 x_1^2 & \lambda_2^2 x_2^2 & \dots & \lambda_n^2 x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^m x_1^m & \lambda_2^m x_2^m & \dots & \lambda_n^m x_n^m \end{pmatrix} = \begin{pmatrix} P^1 \\ P^2 \\ \vdots \\ P^m \end{pmatrix} (X_1, X_2, \dots, X_n). \quad (2)$$

식(2)의 행렬식에서 λ 는 프로젝티브 깊이이며 기본행렬을 통하여 유추하거나 1로 설정한다[5]. 공통의 일치점을 가지는 n 개의 영상에서 식(2)의 좌변이 구해지면 특이값 분해(singular value decomposition)로 프로젝티브 공간상의 카메라정보와 구조정보를 계산해 낼 수 있다. 그러나 공통의 일치점을 가져야 한다는 가정 때문에 프레임 수가 많고 다양한 움직임이 일어나는 일반적인 비디오 영상 등에는 사용할 수가 없다. 이러한 단점을 보완하기 위해서 본 논문에서는 영상을 판별식에 의하여 몇 개의 그룹으로 나누고 각 그룹에 대하여 분해법을 수행함으로써 구성된 그룹을 강건하게 병합하는 과정을 제안하였다.

III. 제안된 알고리즘

영상이 3차원 구조정보를 정확하게 표현하기 위해서는 크게 네 가지 조건을 필요로 한다. 첫째로 충분한 수의 대응점이 필요하고, 둘째로 일치되는 특징점들이 골고루 분포되어야 한다. 또한 일치점들이 한 선이나 면에 집중되어 존재하지 않아야 하며, 마지막으로 자동 카메라 보정을 위해 최소 세장의 영상이 필요하다.

1. 대응점의 개수

분해법을 수행하기 위해 충분한 대응점의 개수가 필요하다. 일반적으로 연속적인 영상 시퀀스에서 첫 번째 프레임과 각 프레임 간의 대응관계는 시간에 따라 줄어들게 되며 이 정보는 분해법을 적용할 수 있는 그룹을 결정하는데 중요한 요소가 된다. 이는 각 그룹의 첫 번째 영상

과 두 번째 영상의 대응점의 개수를 기준으로 시퀀스에 따라 대응점이 줄어드는 비율을 계산함으로써 판별할 수 있다.

$$N_r = \left(1 - \frac{N_m}{N_f}\right), \quad (3)$$

여기서 N_r 는 각 번들의 첫 번째 영상과 두 번째 영상의 대응점 개수, N_m 은 해당영상과 이전 영상과의 대응점의 개수이다.

2. 평면사영변환(planar projective transformation) 오차

평면사영변환행렬에 의한 사영에러는 해당 시퀀스에서 카메라의 이동정도를 나타내며 3차원 구조를 해석하기 위해서는 충분히 커야한다. 동시에 대응점이 3차원 구조상의 여러 면에 골고루 존재하는가를 의미한다. 영상으로부터 정확하게 3차원 구조를 복원하기 위해서는 대응점들이 대상영상 내에 동일한 선이나 면에 집중되어 분포되지 않아야 한다. 회전만 하는 카메라로부터 얻어진 영상사이에는 평면사영변환행렬로 일대일 대응관계를 만족하며, 3차원 구조의 선과 선 그리고 면과 면 사이도 이러한 관계를 만족한다. 이러한 성질은 영상 모자이크 기법 등에 이용된다. 하나의 3차원 선이나 면 위에 존재하는 대응점들은 카메라 자동보정에 부적합하며 이는 평면사영변환 에러가 클수록 프로젝티브 재구성을 하기에 적합하다는 것을 의미한다. 영상과 영상 사이의 평면사영변환행렬은 카메라가 이동함으로써 그리고 대응점이 다양한 선이나 면에 분포함으로써 일대일 대응에 에러를 만들며, 이는 평면사영변환 에러가 영상을 그룹핑할 때 중요한 요소가 된다. 즉 그룹 전체에 대응되는 대응점이 한 선이나 면에만 집중되어 분포한다면 그룹핑을 정지하고 다시 새로운 특징점을 선정하여야 한다.

평면사영변환 에러는 두 영상사이의 최소 4개의 일치정보로부터 구할 수 있으며, 4개 이상의 점에 대해서는 DLT(direct linear transformation)방법을 이용하여 구할 수 있다[6]. 다음은 평면사영변환투영의 에러식이다.

$$H_{err} = \frac{1}{N} \sum_{i=1}^N d(x_i, Hx_i'), \quad (4)$$

여기서 x_i, x_i' 은 대응점으로 동차좌표이고, H 는 평면사영변환행렬, N 은 전체 대응점의 개수이다.

3. 대응점의 분포

영상 시퀀스로부터 3차원 구조정보를 정확하게 재구성하기 위해서는 영상에 대응점이 고르게 분포되어야 한다. 평면사영변환 에러가 대응점의 분포정도를 어느 정도 반영할 수도 있다. 그러나 영상 한쪽에 치우친 물체에만 대응점이 존재하는 경우, 평면사영변환 에러는 클 수 있지만 특징점이 영상 내에 고르게 분포하지 않기 때문에 정확하고 안정적인 프로젝티브 재구성이 어렵다. 일치점이 영상에 고르게 분포하는지를 검사하기 위해 영상을 일정한 크기의 하위 영역으로 분할하여 전체 영상에서의 대응점 밀도와 각 하위 영역에서의 대응점 밀도간의 차이를 이용하여 대상 영상에 대한 일치점의 분포 정도를 해석할 수 있다. 대응점 밀도의 표준편차 식은 다음과 같다.

$$\sigma_p = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{N s_i}{A s} - \frac{N}{A}}^2, \quad (5)$$

여기서 N 은 전체 대응점의 개수이고 $N s_i$ 는 i 번째 하위영역 내에 있는 대응점의 개수, A 는 전체 영역이고 $A s$ 는 각각의 하위영역이다.

4. 제안된 판별식

위에서 정의한 요소들로부터 영상 시퀀스를 그룹화 시키는 판별식은 다음과 같이 정의할 수 있다.

$$S = w_1 N_r + w_2 \frac{1}{H_{err}} + w_3 \sigma_p, \quad (6)$$

여기서 S 는 판별식점수, w_n 은 각 요소의 가중치이다.

식(6)의 판별식 점수가 임계값 이상으로 커지면 그룹을 생성하는 과정을 중지하고 새로운 그룹을 생성한다. 각 그룹을 하나의 프로젝티브 공간으로 위치시키기 위해 각 그룹은 2장의 영상이 공유되도록 한다. 이러한 반복과정을 통해 영상 시퀀스는 몇 개의 그룹으로 나뉘며, 각 그룹의 각기 다른 프로젝티브 공간을 식(7)을 이용하여 평면사영변환행렬(H)을 구하여 하나의 프로젝티브 공간으로 병합할 수 있다.

$$\begin{aligned} P_1' H^{-1} &= P_1 \\ P_2' H^{-1} &= P_2, \end{aligned} \quad (7)$$

여기서 P_1, P_2 는 앞 그룹의 마지막 두 영상의 카메라투영행렬이고 P_1', P_2' 은 뒤 그룹의 처음 두 영상의 카메라 투영행렬이다.

IV. 유클리드 재구성방법

공간사영변환행렬은 프로젝티브 공간을 유클리드 공간으로 변형하는 행렬이다. 공간사영변환행렬은 절대2차원추곡면을 추정하여 분해함으로써 구하며, 절대2차원추곡면은 3장 이상의 영상으로부터 추정된다[2]. 그러나 이러한 선형방정식은 잡음의 영향으로 정확한 해를 얻기 어렵다. [1]에서는 잡음의 영향에 대응하기 위해 오차가 적은 카메라투영행렬을 구분하는 방법을 제시하였다. 본 논문에서는 Ransac(Random Sample Consensus) 알고리즘 대신 몬테카를로 기법을 기반으로 하는 LMeds(least median of square)방법으로 에러가 많은 카메라 행렬을 제거함으로써 정확한 절대2차원추곡면을 추정하였다.

적합한 카메라투영행렬은 정규화 된 카메라 내부 파라미터의 뒤트림요소와 원리점이 0이어야 하고 초점거리의 비율이 1이 되어야 한다. 이러한 성질은 다음과 같이 표현할 수 있다.

$$error = \sum_{j=1}^N (w_1 s^2 + w_2 (1 - \frac{f}{f_j})^2 + w_3 (c_x^2 + c_y^2)), \quad (8)$$

여기서 w 는 가중치, s 는 뒤트림요소, f 는 초점거리 그리고 c 는 원리점이다. 에러가 많은 카메라투영행렬을 절대2차원추곡면을 계산하는 과정에서 제거하기 위해 여러 개의 카메라 투영행렬 중에서 임의로 3개의 카메라 행렬을 선택하여 절대2차원추곡면을 추정하고 추정된 절대2차원추곡면을 나머지 카메라에 투영하여 식(8)을 통해 에러를 계산한다. LMeds방법을 사용해 이러한 과정을 반복하여 최적의 카메라투영행렬의 집합을 구한다.

V. 실험결과

제안된 프로젝티브 재구성 결과를 확인하기 위하여 그림 1의 얼굴조형을 촬영한 21장의 영상을 사용하여 실험하였다. 제안된 방법의 계산시간과 카메라의 오차를 표 1에 보이고 기존의 방법과 비교하였다.

카메라투영행렬의 에러는 식 (8)의 각 요소별로 계산하였고 제안된 방법으로 21장의 영상을 총 4개의 그룹으로 나누었다. 표 1에서 보는 바와 같이 제안된 프로젝티브 재구성 방법을 사용하면 기존의 병합기반 방법에 비해서 계산속도가 현저히 줄어드는 것을 확인할 수 있으며, 계산결과도 크게 차이가 없음을 알 수 있다. 그림 2는 제안된 방법으로 구한 카메라 정보와 병합기반 재구성방법의 정확도를 확인하기 위하여 초기 매칭된 대응점을 이용하여 재구성한 결과이다. 그림 3은 제안된 방법으로 구한 카메라특성 및 위치정보와 스테레오 매칭기법을 통해 계산된 대응점으로 삼각도법에 의해 조밀하게 3차원 재구성한 결과이다.

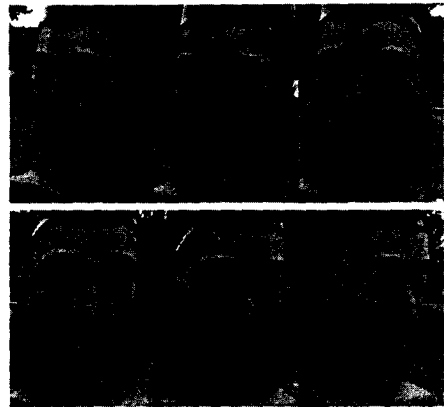
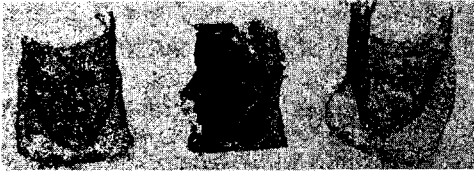


그림 1. 입력영상.

표 1. 제안된 방법과 병합기반 방법의 비교

	계산시간(초)	뷰불일	초점거리 비율	원리점
병합기반	43	2.972	0.996	0.747
제안된 방법	20	2.334	1.204	0.514



a. 평면기반 재구성 결과.



b. 제안된 재구성 결과.

그림 2. 3차원 재구성 결과.

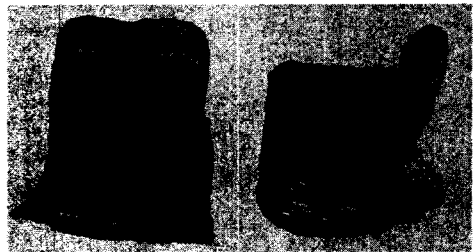
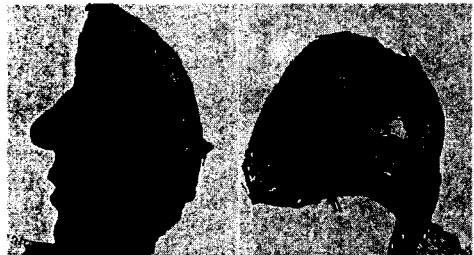


그림 3. 조밀한 3차원 재구성 결과.

VI. 결론

실세계의 물체가 카메라에 맺히는 과정에 있어서 데이터의 정수화로 인해 많은 정보를 잃게 되고, 노이즈와 렌즈의 뒤틀림(lens distortion)현상 등으로 인하여 왜곡된다. 따라서 잡음이 포함된 영상에서 사전정보 없이 3차원 구조정보를 정확하게 복원하는 일은 매우 어렵다. 본 논문에서는 비보정된 영상 시퀀스를 대상으로 강건한 3차원 재구성을 위해 분해법기반 방법을 기반으로 새로운 프로액티브 재구성 알고리즘을 제안하였다. 기존의 분해법기반 방법으로는 해결할 수 없었던 시퀀스 영상에 대해 제안된 방법을 이용하여 적용할 수 있으며, 기존의 정합방법보다 계산속도가 매우 적게 소요되는 것을 확인할 수 있었다.

이후 연구에서는 카메라투영행렬과 영상간의 대응점을 통해 복원된 3차원 구조정보를 최적화하는 연구가 진행될 예정이다.

감사의 글

본 논문은 교육부 두뇌 한국 21(BK21) 사업, 한국과학재단의 해외 Post-doc. 연수지원 및 과학기술부 NRL 사업(2000-N-NL-01-C-285)지원에 의해서 연구되었습니다.

참고문헌

- [1] S.Gibson, J.Cook, T.Howard, R.Hubbold, and D.Oram, "Accurate Camera Calibration for Off-line, Video-Based Augmented Reality," *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, Germany, Sept. 2002.
- [2] B.Triggs, "The Absolute Quadric," *IEEE Conf. CVPR*, pp. 609-614, 1997.
- [3] A.Azarbayejani and A.Pentland, "Recursive Estimation of Motion, Structure, and Focal Length." *IEEE Trans. on Pattern Matching and Machine Intelligence*, vol.17 no. 6, pp.567-575, June 1995.
- [4] A.Chiuso, P.Favaro, H.Jin, and S.Soatto, "Motion and Structure causally Integrated over Time," *IEEE Trans. on Pattern Matching and Machine Intelligence*, vol.24, no. 4, pp.523-535, April 2002.
- [5] P.Sturm and B.Triggs, "A Factorization Based Algorithm For Multi-Image Projective Structure and Motion," *4th European Conference on Computer Vision*. Cambridge, England, pp. 707-720, April 1996.
- [6] R.Hartley and A.Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press. 2000.