

論文 2003-40SP-5-5

2단계 샘플링을 이용한 카메라 움직임 및 장면 구조 복원 (Camera Motion and Structure Recovery Using Two-step Sampling)

徐廷珏*, 趙青雲*, 洪顯基*

(Jung-Kak Seo, Cheung-Woon Jho, and Hyun-Ki Hong)

요약

비디오 영상으로부터 카메라의 움직임과 3차원 구조를 복원하는 기술은 다양한 분야에 응용되고 있다. 특히 비교정(un-calibrated) 동영상은 해석하기 위해서는 대상 영상의 정보만을 이용하는 카메라의 자동 보정(auto calibration)기술이 필수적이다. 그러나 비디오 상의 많은 프레임에 안정적으로 이를 적용하려면 기존의 자동 보정기술은 무리조정(bundle adjustment) 또는 비선형 최적화 등의 매우 복잡한 과정이 요구된다. 본 논문에서는 최적화 과정 없이도 정확하게 대상 카메라의 궤적과 3차원 구조를 복원하는 새로운 방법이 제안된다. 첫 번째 단계에서 대상 시퀀스에서 카메라 궤적의 해석에 적절한 키프레임(key frame)을 선택하여 전체 연산 시간을 줄이며, 두 번째 과정에서 보다 정확한 카메라 자동 보정을 하기 위해 이미 추출된 키프레임 가운데 절대 2차 원추곡면(absolute quadric)의 추정을 통해 오차가 많이 포함된 키프레임을 제거한다. 가상 및 실사영상에 대한 실험결과로부터 제안된 방법의 성능을 확인하였으며, 다양한 실사 영상을 대상으로 가상의 3차원 모델을 합성한 결과도 제시하였다.

Abstract

Camera pose and scene geometry estimation from video sequences is widely used in various areas such as image composition. Structure and motion recovery based on the auto calibration algorithm can insert synthetic 3D objects in real but un modeled scenes and create their views from the camera positions. However, most previous methods require bundle adjustment or non linear minimization process for more precise results. This paper presents a new auto-calibration algorithm for video sequence based on two steps: the one is key frame selection, and the other removes the key frame with inaccurate camera matrix based on an absolute quadric estimation by LMedS. In the experimental results, we have demonstrated that the proposed method can achieve a precise camera pose estimation and scene geometry recovery without bundle adjustment. In addition, virtual objects have been inserted in the real images by using the camera trajectories.

Keyword : structure and motion recovery, auto calibration, key frame, match move

* 正會員, 中央大學校 尖端映像大學院 映像工學科

(Dept. of Image Eng., Graduate School of Advanced Imaging Science Multimedia & Film, Chung-Ang University)

※ 이 논문은 교육부 두뇌 한국 21(BK21) 사업 및 과학기술부 NRL 사업(2000-N-NL-01-C-285) 지원에 의해서 연구되었습니다.

接受日字:2003年6月19日, 수정완료일:2003年9月20日

I. 서론

컴퓨터 비전 기술은 1990년대 이후 시각효과(visual effects)에 널리 응용되고 있으며, 실제 촬영된 필름과 가상 물체를 결합하는 영상합성(image composition) 분야가 대표적이다¹⁾. 2차원 영상 합성은 실사 영상에 가상 영상을 위치시키고 이 위치를 다음 프레임에서 계속 추적하여 합성한다. 그러나 카메라의 움직임이 크고

회전성분이 포함된 경우에는 카메라 정보를 해석하고 장면(scene)에 대한 3차원 모델을 구성하는 3차원 합성 기술이 필요하다. 본 논문에서는 비교정(un-calibrated) 비디오 영상에서 카메라의 움직임 정보와 장면에 대한 3차원 구조의 복원을 목적으로 한다. 두 장 또는 여러 장의 영상에서 카메라 정보를 얻는 연구는 1990년대부터 현재까지 활발하게 진행되고 있다^[2-5]. 최근에는 카메라의 자동 보정기술을 이용하여 비디오 영상으로부터 3차원 구조를 모델링하거나 카메라 궤적을 추정하여 가상 물체를 합성하는 등의 증강 현실관련 연구가 진행되고 있다^[2, 6-8]. H. S. Sawhney 등은 비디오 영상으로부터 카메라 궤적을 얻기 위해 기본행렬(fundamental matrix)을 이용하였으며, 기본행렬로부터 카메라의 외부 파라미터를 추출하여 3차원 매치무브(match move)를 구현했다^[6]. M. Pollefeys 등은 영상 시퀀스 내의 구조를 3차원적으로 모델링하는 기법을 제안했으며, 이를 확장하여 비디오 영상에서 증강현실(augmented reality)을 구현하였다^[8]. S. Gibson 등은 특징점 추적 기법을 향상시키고, 계층적 정합기반의 사영구조를 복원하였으며, 카메라의 자동보정과정에 RANSAC 기법을 도입하여 전체 카메라 궤적을 추정하였다^[2]. 비디오 시퀀스가 매우 많은 프레임으로 구성되어 있기 때문에 기존 연구들은 많은 계산량이 필요한 무리조정(bundle adjustment) 등의 최적화 과정이 요구된다.

본 논문에서는 최적화 과정 없이도 비디오 영상에서 정확하게 카메라 궤적과 3차원 구조를 복원하는 새로운 방법이 제안된다. 먼저 대상 비디오 영상으로부터 카메라의 움직임 해석에 적합한 키프레임(key-frame)을 선택하여 전체 연산 시간을 줄이며, 두 번째 과정에서는 카메라 자동 보정(auto-calibration)을 하기 위해 절대 2차 원추곡면(absolute quadric)을 추정하며, 이 과정에서 에러가 많이 포함된 키프레임을 추가로 제거한다. 제안된 방법에서는 키프레임 선택과정에서 인라이어(inlier)의 분포를 고려함으로써 처리 시간을 단축할 수 있다. 또한 카메라의 자동보정 과정에 LMedS (least median of square)기술을 기반으로 하여 절대 2차 원추곡면을 추정한다. 절대 2차 원추곡면의 추정은 프로젝티브 카메라 행렬을 통해 얻어지며, 기존 방법에서는 프로젝티브 카메라를 모두 사용하여 절대 2차 원추곡면을 추정한다. 그러나 잡음이 많이 포함된 프로젝티브 카메라행렬이 존재하면, 이로 인해 무리조정의 결

과가 많이 영향을 받게 된다. 본 논문에서는 LMedS를 이용하여 잡음이 많은 카메라 정보는 제거하며, 잡음 적은 카메라 정보만으로 절대 2차 곡면을 추정한다. 여기서 키프레임으로 선택되지 않은 영상들의 카메라 정보는 복원된 3차원 구조 정보와 영상 내 일치점 정보를 기반으로 카메라 절제(resection)을 통해 복원된다^[9].

II장에서는 카메라행렬과 자동 보정에 대해 서술하며, III장에서는 키프레임 선택과정 및 개선된 키프레임선택 판별식 및 LMedS방법을 이용하여 절대 2차 원추곡면의 추정방법이 제안된다. 그리고 IV장에서 실험 및 검토를 통해 제안된 방법의 우수한 성능을 확인한 다음, V장에서 결론을 맺는다.

II. 자동 카메라 보정(Auto-calibration)

3차원 공간은 카메라에 의해 2차원 평면에 사영되며, 이 관계는 식 (1)과 같이 3×4 투영행렬(projection matrix)로 표현된다.

$$P = K(I | 0) \begin{pmatrix} R_{3 \times 3} & t_3 \\ 0 & 1 \end{pmatrix}, \quad (1)$$

여기서 투영행렬은 3개의 행렬곱이며, $R_{3 \times 3}$ 및 t_3^T 는 카메라의 외부(extrinsic) 파라미터로 강제 변환(rigid transformation) 행렬에서 회전 및 이동 성분이다. $(I | 0)$ 항이 3차원 공간을 영상으로 투영하며, K 는 내부(intrinsic) 파라미터로 평면에 투영된 것을 영상 내의 좌표계로 변환한다.

월드 좌표계는 사용자가 설정하며, 카메라의 위치와 시선방향을 월드 좌표계에 일치시키면, 첫 번째 카메라 행렬은 $P_1 = K[I | 0]$ 가 된다. 같은 카메라를 회전 및 이동하면, 두 번째 영상의 행렬은 $P_2 = K[R | t]$ 로 표현된다. 비보정 영상을 대상으로 하기 때문에 K 는 미지이며, 첫 번째 카메라 행렬에 K 를 소거하도록 4×4 변환행렬을 적용하여 다음 식 (2)를 얻는다.

$$\begin{aligned} x_1 &= P_1 X = K[I | 0] X = K[I | 0] H H^{-1} X = [I | 0] X, \\ x_2 &= P_2 X = K[R | t] X = K[R | t] H H^{-1} X = [M | p_1] X, \end{aligned} \quad (2)$$

여기서 X 는 3차원 공간의 점이고, x_1, x_2 는 두 영상에 투영된 점, X_p 는 X 의 사영변환된 프로젝티브 구조다. 두 장에서 기본행렬을 이용해 프로젝티브 구조를 복원하며, 프로젝티브 구조로부터 공간 사영변환행렬 H 를

구하면 유클리드 구조의 복원도 가능하다^[9]. 또한 카메라 투영행렬도 복원할 수 있기 때문에 내부 및 외부 파라미터의 추정도 가능하다(그림 1). 따라서 카메라의 파라미터를 추정하기 위해서는 프로젝티브 구조 복원과 이로부터 공간사영변환을 정확하게 추출하는 것이 중요하다.

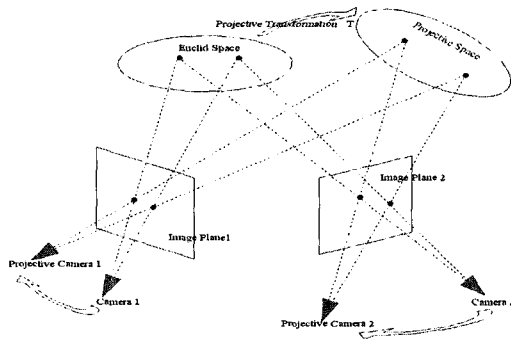


그림 1. 유클리드 구조와 프로젝티브 구조 관계
Fig. 1. Relation of Euclidean and projective framework.

1. 프로젝티브 구조 복원

두 영상에서는 기본행렬을 이용하여 프로젝티브 구조의 복원이 가능하며, 세 장의 영상에서는 3초점 텐서(trifocal tensor)로 복원한다^[9]. 영상 시퀀스로부터 프로젝티브 구조를 복원하는 방법은 크게 분해법(factorization) 및 병합(merging) 방법으로 구분된다^[2]. 분해법 방법은 카메라 투영행렬과 3차원 구조가 동시에 얻어지기 때문에 오차 누적에 의한 영향을 보다 적게 받는다. 이에 비해 병합기반 방법은 초기 추정의 정확성에 매우 영향 받는 특징이 있으며, 비디오 영상 등의 시퀀스에 적용하기 적합하다^[10, 11]. 정합기반 방법은 순차적 병합과 계층적 병합방법으로 구분되며, 순차적 방법이 초기 추정치에 매우 의존하고 에러가 누적되는 반면, 계층적인 방법은 이를 줄일 수 있는 장점이 있다.

2. 절대 2차 원추곡면 추정

절대 2차 원추곡면은 무한면이 절단하며, 이는 곡선 형태를 띠고 절대 원추곡선(absolute conic)이라고 한다. 절대 2차 원추곡면은 4x4 대칭행렬로 표현되며, 계수(rank)가 3이다. 또한 유클리드 공간에서는 $\begin{pmatrix} I_{3 \times 3} & 0 \\ 0 & I \end{pmatrix}$ 의 형태가 된다. 유클리드 상의 카메라 행렬 P와 프로젝티브 구조의 카메라 행렬 P_p의 관계는 임의 공간사

영변환 H로 설명된다. 식 (3)에서 카메라의 내부 파라미터 중에서 화소의 가로, 세로비가 1, 카메라의 비틀림(skew)을 0, 원리점(principle point)을 영상의 중앙이라고 가정하면, 식 (4)와 같은 선형 방정식을 유도할 수 있다.

$$PQT^T = KK^T. \tag{3}$$

$$(P_p Q_p T_p^T)_{12} = (P_p Q_p T_p^T)_{13} = (P_p Q_p T_p^T)_{23} = 0,$$

$$(P_p Q_p T_p^T)_{11} = (P_p Q_p T_p^T)_{22}. \tag{4}$$

여기서 (A)_{i,j}는 A행렬의 i번째 행, j번째 열의 원소를 의미하며, Q_p는 4x4 대칭 행렬로 프로젝티브 공간에서의 절대 2차 원추곡면이고, 10개의 미지수와 9 자유도를 가진다. 프로젝티브 구조에서의 카메라 행렬로부터 4개의 방정식을 만들 수 있으며, 최소 4개의 카메라 행렬로부터 Q_p를 추정할 수 있다. 유클리드 공간에서의 Q와 프로젝티브 구조의 Q_p 관계에서 Q_p를 추정한 후에 고유치분해(eigenvalue decomposition)를 통해 H⁻¹를 구한다.

$$EVD(Q_p) = UDU^T = U\sqrt{D} \begin{pmatrix} I_{3 \times 3} & 0 \\ 0 & I \end{pmatrix} \sqrt{D}, \tag{5}$$

여기서 D는 대각행렬로 각 원소에 제곱근을 하여 쉽게 분리할 수 있으며, U 행렬과 곱하여 H⁻¹를 얻는다. 절대 2차 원추곡면은 계수가 3이므로 \sqrt{D} 의 마지막 원소는 0이 된다. 따라서 H⁻¹를 비가역행렬을 만들며, \sqrt{D} 의 마지막 원소에 임의의 숫자를 넣어 H⁻¹를 얻고 보통 1로 대체한다^[6].

III. 카메라 움직임 및 3차원 구조 복원

본 장에서는 최적화 과정 없이도 정확하게 카메라 제적과 3차원 구조를 복원하는 새로운 방법이 제안된다. 이를 위해 카메라 해석 등에 적합한 프레임 집합을 구성하며, 먼저 대상 비디오 영상에서 키프레임을 선택한다. 그리고 보다 정확하게 카메라를 자동 보정하기 위해 절대 2차 원추곡면의 추정과정 중에 에러가 많이 포함된 키프레임을 LMedS 방법을 이용하여 제거한다. 두 과정에서 제외된 카메라 정보는 최종적으로 카메라 절제를 통해 복원된다.

1. 1단계 샘플링: 키 프레임 선택

비교적 비디오의 많은 프레임 중에서 카메라 움직임

및 장면구조를 복원하기에 적합한 키프레임을 효과적으로 선택함으로써 자동 보정과정의 전체 계산량을 대폭 줄일 수 있으며, 카메라 움직임이 균등하게 분포되는 장점이 있다. 즉, 보다 효율적으로 카메라를 보정하기 위해 가능한 적은 수의 키프레임이 선택되어야 한다. 또한 특징점의 추적 등이 프레임 간에 효과적으로 진행될 수 있도록 키프레임이 선택되어야 하며, 선택된 영상들에서 카메라의 움직임 정보를 정확하게 해석할 수 있어야 한다. 본 논문에서는 키프레임을 선택하기 위한 새로운 방법이 제안된다.

H. S. Sawhney 등은 키프레임을 효과적으로 선택하기 위해, (i) 카메라 움직임이 일정할 때 시간에 대해 규칙적으로 선택하며, (ii) 카메라의 패럴랙스(parallax) 움직임 사이에서는 설정된 기준의 임계값 보다 적은 영상, (iii) 일치점의 수가 급격히 변하는 영상, (iv) 사용자의 의도 등을 고려하였다^[6]. D. Nister는 대상 영상의 첨예도(sharpness)와 평면사영변환의 투영오차를 이용하는 키프레임 선택기법을 제안했다^[7]. 이 방법은 첨예도가 낮은 영상부터 제거하여 영상 내에서 특징점 추출이 어려운 영상이 키프레임으로 선택되는 것을 피하고, 평면사영변환 투영오차를 연결성(connectivity)으로 이용하여 일치 정보가 충분히 카메라 움직임 정보를 표현하는지를 검사한다. 그러나 매 프레임마다 영상의 첨예도를 계산해야 하며, 다른 방법에 비해 상대적으로 많은 키프레임이 선택된다. S. Gibson 등은 일치점의 수, 평면사영 변환투영 에러, 기본행렬 추정시의 중간값 에러를 조합하여 키프레임 선택을 위한 판별식을 제시하였다^[2]. 이 방법은 매 프레임마다 기본행렬을 추정하며, 이 과정은 첨예도 검사보다 많은 계산이 요구된다.

1.1 키 프레임 선택 판별식

일반적으로 선택된 영상으로부터 카메라의 움직임을 해석하려면, 대상 영상간의 일치 관계를 효과적으로 구성해야 하며, 대상 장면과 카메라 사이의 거리가 클수록 카메라의 움직임을 정확하게 해석할 수 있다^[7]. 이 조건을 만족하는 영상을 선택하기 위해 제안된 방법은, (i) 특징점 수에 대한 일치점 수의 비, (ii) 두 영상 사이의 평면사영변환행렬에 의한 사영에러, (iii) 영상에 대한 일치점의 분포 등을 고려한다.

특징점 대 일치점 수의 비는 시퀀스 간에 존재하는 일치점이 충분한가를 분석한다. 비디오 영상에서 카메라의 움직임에 의한 영상의 변화는 일반적으로 시간에

따라 일치점의 수가 줄어들며, 이 정보는 다음 키프레임이 선택될 범위의 결정에 적합하다. 평면사영변환행렬에 의한 사영에러는 현 프레임에서 카메라의 이동 정도를 의미한다. 그리고 대상 영상에서의 일치점 분포는 선택된 프레임에서 정확한 에피폴라 기하(epipolar geometry)의 추정 가능성을 검사하기 위해 고려된다. 두 영상사이의 카메라 움직임 정보를 포함하는 에피폴라 기하는 대수적으로 기본행렬로 표현되고, 이는 영상간의 일치점 정보로부터 추정된다. 기본행렬의 추정은 잡음에 매우 민감하며, 일치점이 영상에 고르게 분포할수록 보다 정확한 기본행렬을 추정할 수 있다^[8,13]. 식 (6)은 효과적인 키프레임을 선택하기 위해 위의 요소들을 고려한 판별식이다.

$$S_{ij} = \omega_1 \left(1 - \frac{N_C}{N_F}\right) + \omega_2 \sigma_p + \omega_3 \left(\frac{1}{H_{err}}\right), \quad (6)$$

여기서 S_{ij} 는 i 및 j 프레임간의 판별식 점수, ω 는 가중치, 그리고 N_C , N_F 는 일치점, 특징점의 개수, σ_p 는 대상 영상에서 일치점의 분포 표준편차, H_{err} 는 사영변환행렬에 의한 사영에러다. 가중치는 식 (6)에서 각 항의 가중치를 의미하며, 여러 실험을 통해서 $\omega_1=3$, $\omega_2=1$, $\omega_3=10$ 로 결정했다.

제안된 판별식을 이용해 키프레임을 선택하며, 시퀀스의 첫 번째 및 마지막 프레임도 키프레임으로 설정한다. 첫 번째 영상의 특징점 수에 대해 다음 영상의 특징점 수의 비율이 50%가 될 때까지 판별식 점수를 계산하며, 점수가 가장 작은 영상을 키프레임으로 선택한다. 마지막 프레임까지 위의 과정을 반복하면서 키프레임을 계속 얻는다.

1.2 평면 사영변환 에러

이동하지 않고 회전만 하는 카메라로부터 얻어진 영상 시퀀스에는 평면사영변환행렬로 일대일 대응관계가 존재하며, 이는 영상 모자이크 기법 등에 이용된다. 카메라의 이동이 없는 영상에서 3차원 구조 정보를 복원할 수 없으며, 따라서 평면사영변환 오차가 작을수록 카메라 자동보정에는 부적합하다는 것을 의미한다^[9]. 영상간의 평면사영변환행렬은 카메라의 이동에 따라 일대일 관계가 성립하며, 크게 이동할수록 깊이 정보에 따른 사영오차는 커진다. 즉, 사영오차는 두 카메라의 상대적인 움직임 정보를 나타내며, 이는 카메라 사이의 거리를 나타낸다. 사영오차는 영상 간에 최소 4개의 일치정보로부터 구할 수 있으며, 4개 이상의 점에 대해서

는 DLT(direct liner transformation algorithm) 방법을 이용해 구한다^[9]. 다음 식 (7)은 평면사영변환투영의 에러식이다.

$$H_{err} = \frac{1}{N_C} \sum_{i=1}^{N_C} d(x_i, Hx'_i), \quad (7)$$

여기서 x_i, x'_i 는 일치점으로 동차좌표, H 는 평면사영변환행렬이다.

1.3 일치점 분포 표준 편차

일치점들의 분포는 기본행렬의 추정에 많은 영향을 주며, 영상에 대해 일치점 분포가 고른 경우에 더 정확하게 기본행렬을 추정할 수 있다^[9]. 일치점이 영상에 고르게 분포하는지를 검사하기 위해 본 논문에서는 영상을 인라이어의 수에 따라 일정한 크기의 부영역(sub region)으로 분할하여 전체 영상에서의 일치점 밀도와 각 부 영역에서의 일치점 밀도간의 차이를 이용하여 대상 영상에 대한 일치점의 분포 정도를 해석한다. 만일 나누어진 영역의 개수가 일치점의 개수와 같은 경우에서 각 영역에 대해 점이 한 개씩 존재한다면, 일치점들이 영상 내에 가장 고르게 분포되었다고 볼 수 있다. $W \times H$ 크기의 대상 영상을 균일한 크기의 부영역으로 나눌 때 그 형태를 전체 영상의 형태와 같은 사각형으로 하는 것이 유리하며, 전체 영상을 분할하기 위해 다음 식 (8)을 이용한다^[13].

$$W_S = W / \text{int}(\sqrt{N_C}), \quad H_S = H / \text{int}(\sqrt{N_C}), \quad (8)$$

여기서 $\text{int}(\cdot)$ 는 정수형 변환을 의미하며, W_S 및 H_S 는 각각 부영역의 가로, 세로 크기이다. 식 (8)을 이용하여 부영역의 크기를 구하고, 부영역의 각 방향의 크기로 전체 대상영상을 분할한다.

대상영상에 대한 일치점의 분포 정도를 계산하기 위해 각 부영역 안에 존재하는 일치점 수를 구하고 전체 영상의 면적에 대한 일치점 밀도와 각 부영역 내에 존재하는 일치점의 밀도에 대한 표준편차를 이용한다. 부영역은 크기가 일정하기 때문에 그 면적을 1이라 하면, 점 밀도에 대한 표준편차 식은 다음과 같다.

$$\sigma_p = \sqrt{\frac{1}{N_S} \sum_{i=1}^{N_S} (N_{ci} - \frac{N_C}{N_S})^2}, \quad (9)$$

여기서 N_S 는 나누어진 부영역의 개수이고, N_{ci} 는 i 번째 부영역 안에 있는 일치점의 개수다.

2. LMedS 기반의 절대 2차 원추곡면 추정

세 장 이상의 영상으로부터 절대 2차 원추곡면은 선형방정식을 통해 추정될 수 있지만, 이러한 선형방정식은 잡음의 영향으로 인해 정확한 해를 얻기 어렵다. 본 절에서는 몬테카를로 기법을 기반으로 하는 LMedS 방법으로 에러가 많은 카메라 행렬을 제거함으로써 보다 정확한 절대 2차 원추곡면을 추정하는 새로운 방법이 제안된다.

프로젝티브 카메라 행렬 중에 첫 번째 카메라 행렬이 $[I \mid 0]$ 가 되도록 모든 카메라 행렬에 공간사영행렬을 곱한다. $P = [M \mid p_4]$ 라고 하면, 다음의 공간사영행렬을 적용하여 쉽게 변환할 수 있다.

$$T = \begin{pmatrix} M^{-1} & -M^{-1}p_4 \\ 0 & 1 \end{pmatrix}. \quad (10)$$

프로젝티브 구조상의 절대 2차 원추곡면을 첫 번째 카메라에 투영하면 절대 2차 원추곡면의 미지수가 10개에서 7개로 감소하며, 이는 첫 번째 카메라를 제외한 최소 2개의 카메라 행렬로부터 추정될 수 있다. 이 과정에서 에러가 많은 카메라 행렬을 제거하기 위해, 첫 번째 카메라 행렬을 제외한 카메라 행렬집합에서 임의로 2개의 카메라 행렬을 선택하여 절대 2차 원추곡면을 추정한다. 추정된 원추곡면을 다시 카메라에 투영하여 첫 번째 카메라 행렬에 투영한 것과 각 카메라에서의 오차(식 11)를 계산하고 순서화하여 중간위치의 오차를 저장한다.

$$r_i = (\lambda_0 P^0 Q_p P^{0'} - \lambda_i P^i Q_p P^{i'})_{norm}, \quad (11)$$

여기서 $norm$ 은 크기, $(KK^T)_{33}=1$ 이므로 λ_0 와 λ_i 는 $1/(P_p^0 Q_p P_p^{0'})_{33}$ 이 된다. 위 과정을 임의 횟수만큼 반복하며, 중간위치의 오차가 가장 작은 경우의 절대 2차 원추곡면을 저장한다. 최소 중간값과 그 때의 2차 원추곡면을 이용하며 에러가 많은 카메라 행렬을 구별하는 임계값을 식 (12)로 구한다.

$$\tau = 2.5 \times 1.4826 [1 + 5/(p-q)] \sqrt{median}, \quad (12)$$

여기서 $median$ 은 최소 중간값이며, p = 카메라 행렬 수 - 1, q 는 임의로 선택하는 카메라 행렬의 수인 2다. 식 (12)로부터 임계값을 정하고, 임계값 보다 큰 오차를 갖는 카메라 행렬을 제외시킨 다음, 다시 절대 2차 원추곡면을 추정한다.

다음은 알고리즘을 정리한 것이다.

- ① 프로젝티브 복원을 한다.
- ② 프로젝티브 카메라 행렬을 첫 번째 카메라 행렬이 $[I | 0]$ 가 되도록 식 (10)을 이용하여 변환한다.
- ③ 첫 번째 카메라 행렬을 제외한 나머지 카메라 행렬에서 임의의 카메라 행렬 2개를 선택한다.
- ④ 선택된 2개의 카메라 행렬로 식 (4)를 이용하여 절대 2차 원추곡면을 추정한다.
- ⑤ 추정된 절대 2차 원추곡면으로 식 (11)를 이용하여 각 카메라 행렬의 오차를 계산한다.
- ⑥ 계산된 오차를 순서화하여 중간위치의 오차값과 절대 2차 원추곡면을 저장한다.
- ⑦ 3-6번 과정을 임의의 횟수만큼 반복하며 중간오차가 가장 작은 경우의 절대 2차 원추곡면과 중간오차값을 저장한다.
- ⑧ 최소 중간오차값으로 식 (12)으로부터 임계값을 정하고, 식 (11)을 이용하여 에러가 많은 카메라 행렬을 제거한다.
- ⑨ 나머지 카메라 행렬을 이용하여 보다 정확한 절대 2차 원추곡면을 다시 추정한다.

위의 방법에 의해 추정된 절대 2차 원추곡면을 이용하여 공간사영변화행렬을 구하고, 이를 프로젝티브 구조와 카메라 행렬에 적용함으로써 대상 영상에 대한 3차원 유클리드 구조를 복원한다.

2단계 샘플링 과정에서 제외된 프레임의 카메라 행렬은 선택된 키프레임에 의해 재구성된 3차원 구조 및 2차원 영상간의 일치 정보를 이용하여 복원된다. 최소 6개의 일치 정보로부터 카메라 행렬을 추정하는 카메라 절제 과정을 이용한다^[9]. 먼저 3차원 구조 정보와 일치하는 2차원 정보를 얻기 위해 인접 카메라 행렬과 스테레오 정합 기법을 이용해 일치점을 구성한다. 스테레오 정합 과정에서 잡음 등에 대처하도록 에피폴라 조건을 통해 일치점 목록을 만들며, RANSAC 기술을 적용한다^[13].

IV. 실험 결과 및 고찰

1. 키프레임 선택 실험

본 논문에서는 세 개의 비디오 시퀀스에 대해 실험하였으며, 기존 방법 및 제안된 알고리즘에 의해 선택

된 키프레임 수와 처리 시간을 <표 1>에 비교하였다. 실험환경은 CPU-Pentium4 2.3GHz, RAM 1Gbytes다. Nister 방법(A)은 매 프레임에 대해 첨예성을 분석하기 때문에 영상 크기에 따라 많은 계산 시간이 소요된다. Gibson 방법(B)은 Nister의 방법에 비해 일반적으로 더 적은 키프레임을 선택하지만, 기본행렬 추정에 많은 연산시간이 요구된다. 결국 <표 1>에서 Gibson 방법이 Nister에 비해 키프레임을 적게 선택하는데 비해서 총 처리시간에는 큰 차이가 없다. 제안된 방법(C)은 기본행렬을 직접 추정하여 이용하는 대신에 일치점 분포를 통해 정확한 기본행렬이 추정될 수 있는지를 사전에 분석한다. 이는 처리시간을 매우 단축시킬 수 있으며, 최종적으로 선택된 키프레임의 수와 위치가 Gibson 방법과 거의 유사하다.

표 1. 추출된 키프레임수와 처리시간 비교
Table. 1. Comparison of the number of the selected key frames and computation time.

실험영상			키프레임 수			처리 시간(sec.)		
영상	프레임수	크기	A	B	C	A	B	C
박스	621	720×480	23	14	14	412	416	179
책상	407	720×480	17	13	13	201	200	76
분수대	134	320×240	28	12	12	16.63	38.59	10
시골집	100	720×576	27	27	24	74	102	42

2. 가상 영상에 대한 성능 실험

제안된 방법의 성능을 효과적으로 평가하기 위해 가상 영상에 대해 실험하였다. 3차원 가상 물체는 <그림 2>와 같이 구성하였으며, 카메라 궤적은 물체를 중심으로 회전하면서 y축으로 상승하도록 하였다. 카메라의 내부 파라미터를 임의로 고정했으며, 영상에 -0.5 ~ 0.5의 잡음을 무작위로 추가하였다.

가상 데이터에 대해 순차적인 프로젝티브 복원방법과 계층적인 방법을 수행했으며, 각 방법에 따라 선형 방법, 무리조정방법, 제안된 방법으로 절대 2차 곡면을 추정하였다. <그림 3>의 복원된 카메라 궤적에서 두 가지 프로젝티브 복원 방법 모두 선형방법에 비해 무리조정 방법이 정확함을 보이며, 제안된 방법은 계산과정이 복잡한 무리조정 방법을 고려하지 않아도 실제 카메라 궤적과 유사한 결과를 얻을 수 있다. 절대 2차 곡면을 이용하여 카메라 파라미터를 얻고 각 카메라 파라미터에 대한 오차를 누적하여 <그림 4>에 나타내

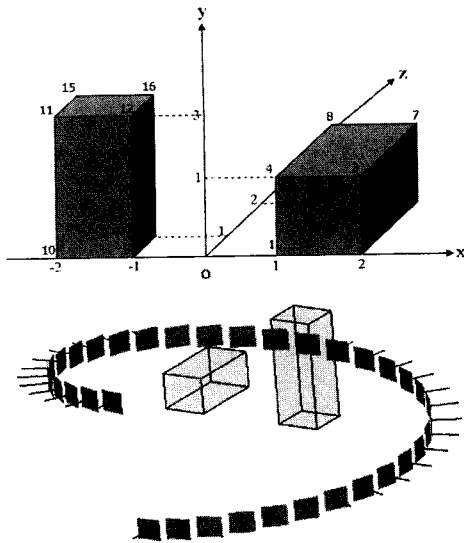


그림 2. 가상 물체와 카메라 궤적
Fig. 2. Synthetic model and camera poses.

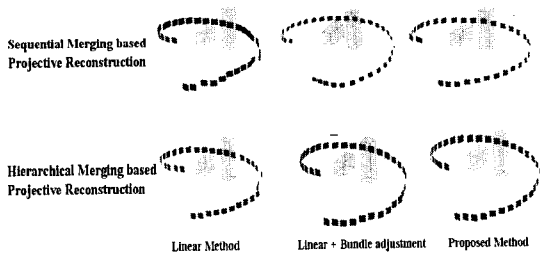
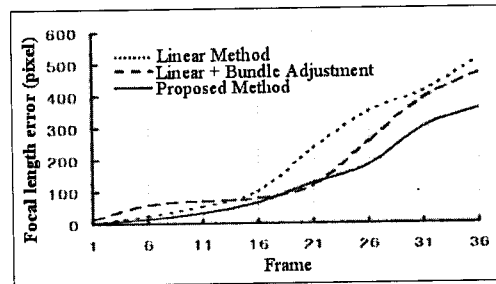
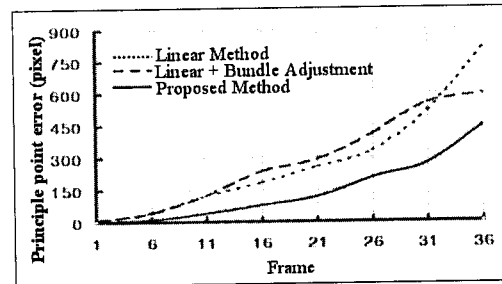


그림 3. 순차 및 계층적 프로젝티브 복원방법에 의한 카메라 궤적의 복원 결과 비교
Fig. 3. Comparison of camera poses recovery on sequential and hierarchical merging method based on projective reconstruction.

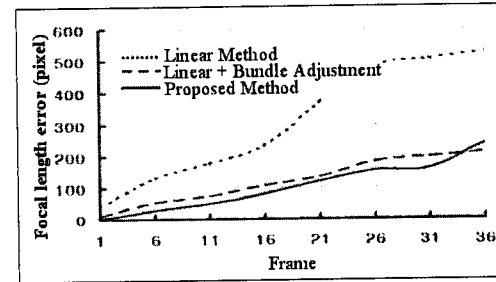
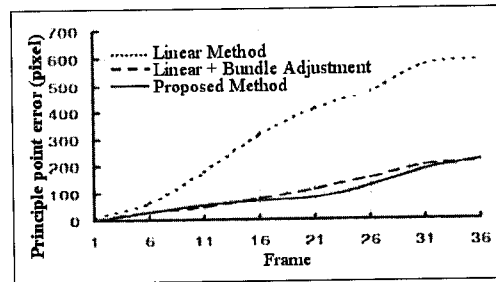
었다. 카메라 궤적의 복원실험과 마찬가지로 선형방법에 비해 무리조정이 우수함을 보이며, 순차적인 프로젝티브 구조 복원 방법에 비해 계층적인 방법이 상대적으로 우수한 성능을 보임을 알 수 있다. 또한 제안된 방법은 두 가지 프로젝티브 구조 복원 과정에서 모두 우수하며, 순차적인 프로젝티브 복원 방법에서는 무리조정 방법에 비해 성능이 매우 향상되었음을 확인하였다.

3. 실사영상에 대한 성능 실험

실사영상에 대해 기존 방법과 제안된 알고리즘의 성능을 비교 분석하였다. 박스 영상의 총 프레임은 621장이고, 크기는 720×480이다. <그림 5(a)>는 대상 비디오



Sequential Merging based Projective Reconstruction



Hierarchical Merging based Projective Reconstruction

그림 4. 순차 및 계층적 프로젝티브 복원방법에 의한 카메라 내부 파라미터의 오차 누적 그래프

Fig. 4. Accumulation error graph for camera intrinsic parameters by sequential and hierarchical merging method based on projective reconstruction.

프레임 3장이며, 제안된 키프레임 판별식에 의해 선택된 영상을 카메라 움직임 그래프 <그림 5(b)>에 도시했다. 움직임 그래프는 매 프레임에서 인접 영상간의

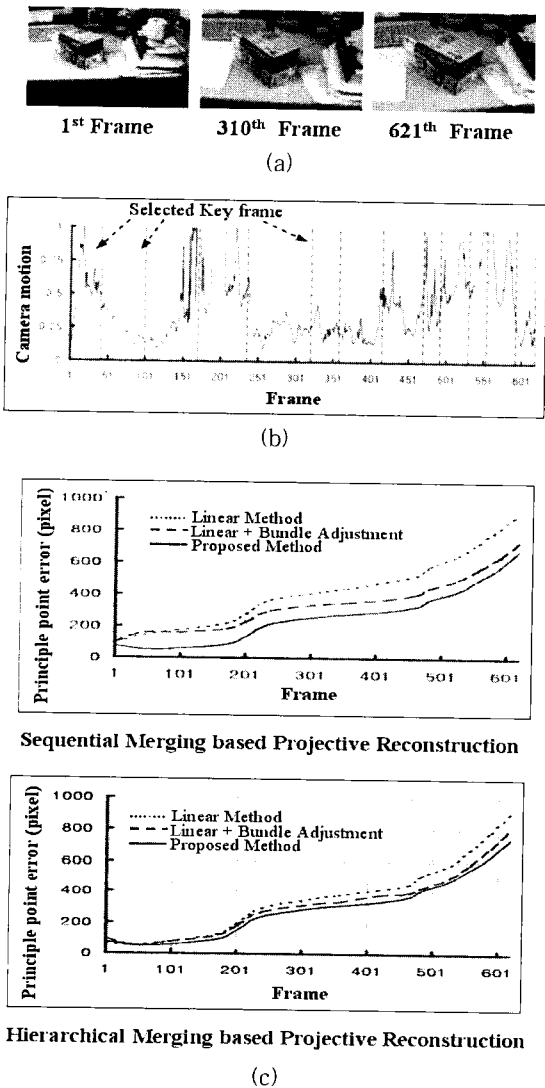


그림 5. 실사영상에 대한 실험. (a) 박스영상(1, 310, 621번째 프레임) (b) 카메라 움직임 그래프와 선택된 키프레임. (c) 기존 방법과 제안된 방법을 순차적 및 계층적 프로젝티브 복원 방법에 적용한 원리점 거리누적오차 그래프

Fig. 5. Experiment results of real video sequence. (a) video frame (1st, 310th, 621th) (b) camera motion graph and selected key frames (c) Accumulation error graph for principal point of camera by sequential and hierarchical merging method based on projective reconstruction.

평면 사영변환 오차를 나타내었다. 전체 시퀀스 상에서 14장의 키프레임이 선택되었고, 시간은 2분 59초가 걸렸다(표 1). 점선으로 표시된 부분이 키 프레임으로 선택된 영상을 나타내며, 그래프는 움직임의 양이 클수



그림 6. 가상의 3차원 물체를 합성한 영상 시퀀스
Fig. 6. Augmented video sequences.

록 키프레임이 상대적으로 많이 선택된다. <그림 5(c)>는 순차적 및 계층적인 프로젝티브 복원 방법에 대해 선형방법, 무리조정 방법, 제안된 방법을 적용하여 얻은 결과로, 영상 내 원리점과의 거리 오차를 누적하여 그래프에 나타내었다. 순차적인 프로젝티브 복원 및 계층적 복원 방법 모두에서 상대적으로 제안된 방법의 우수한 결과를 확인할 수 있다. 또한 <그림 5(c)>로부터

계층적 프로젝티브 구조 복원방법의 특징인 에러분산이 가상영상에서는 잘 반영되지만, 실사영상에서는 다른 요인으로 인해 그렇지 못함을 확인하였다. <그림 6>은 제안된 알고리즘을 여러 비디오 시퀀스 상에 적용하여 얻어진 3차원 구조와 카메라 궤적을 이용하여 가상의 물체를 실사영상에 합성한 결과다.

V. 결 론

비디오 시퀀스 상에 가상 3차원 물체를 정확하게 합성하기 위해 2단계 샘플링을 이용한 카메라 움직임 및 3차원 장면구조 복원방법이 제안되었다. 기존 방법들은 최적의 해를 구하기 위해 무리 조정 등의 최적화 과정을 필요로 하며, 이는 프레임 수가 많을수록 복잡하고 어려운 문제가 된다. 이런 문제를 해결하기 위해 본 논문에서는 효과적인 키프레임 집합을 구성한다. 먼저, 특징점 수에 대한 일치점의 수의 비율, 평면사영행렬 에러, 그리고 영상에 대한 일치점의 분포를 고려한 키프레임 선택 판별식을 통해 키프레임을 선택한다. 또한 프로젝티브 구조를 복원한 후에, LMedS 방법을 기반으로 하여 절대 2차 원추곡면을 추정하는 과정에서 부정확한 키프레임을 추가로 제거한다. 제안된 방법의 성능을 검증하기 위해 가상 및 실사 영상에 대해 실험했으며, 기존 방법과의 비교를 통해 우수한 성능을 확인하였다. 여러 영상에 대해 복원된 카메라 궤적을 이용하여 가상의 3차원 물체를 비디오 시퀀스 상에 합성한 결과도 보였다.

참 고 문 헌

- [1] D. Roble, "Vision in film and special effects," SIGGRAPH Computer Graphics Newsletter, vol. 33, no. 4, pp. 58-60, Nov. 1999.
- [2] S. Gibson, J. Cook, T. Howard, R. Hubbard, and D. Oram, "Accurate camera calibration for off-line, video-based augmented reality," In proc. IEEE and ACM ISMAR, Darmstadt, Germany, Sep. 2002.
- [3] O. Faugeras, Q-T. Luong, and S. Maybank, "Camera self-calibration : Theory and experiments," In proc. ECCV'92, Lecture Notes in Computer Science, vol. 588, pp.321-334, 1992.
- [4] A. Zisserman, P. Beardsley, and I. Reid, "Metric calibration of a stereo rig," In proc. IEEE Workshop Representation of Visual Scenes, pp.93-100, Cambridge, 1995.
- [5] B. Triggs, "Autocalibration and the absolute quadric," In proc. IEEE CVPR, pp. 609-614, 1997.
- [6] H. S. Sawhney, Y. Guo, J. Asmuth, and R. Kumar, "Multi-view 3D estimation and applications to match move," In proc. IEEE MVIEW, pp. 21-28, 1999.
- [7] D. Nister, "Frame decimation for structure and motion," In proc. SMILE 2000, Dublin, Ireland, July 2000.
- [8] K. Cornelis, M. Pollefeys, M. Vergauwen, and L. V. Gool, "Augmented reality from uncalibrated video sequences," In proc. SMILE 2000, Lecture Notes in Computer Science, vol. 2018, pp. 144-160, 2001.
- [9] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge Univ. Press. 2000.
- [10] A. Azarbayejani and A. Pentland, "Recursive Estimation of Motion, Structure, and Focal Length," In proc. IEEE Trans. on Pattern Matching and Machine Intelligence, vol. 17, no. 6, pp. 567-575, June 1995.
- [11] 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.
- [12] R. Hartley, "In Defence of the 8-Point algorithm," In proc. IEEE ICCV, pp. 1064-1070, 1995.
- [13] J. Seo, J. Jung, C. Jho, and H. Hong, "Detection of The inlier point set for a precise fundamental matrix estimation," In proc. FCV, pp.135-140, Jeju, Feb. 2003.

 저 자 소 개

徐廷珏(正會員)

2001년 : 대전대학교 건축공학과 공학사. 2003년 : 중앙대학교 첨단영상대학원 영상공학과 공학석사. <주관심분야 : 컴퓨터 비전, 영상처리, 컴퓨터그래픽스 등>

趙靑雲(正會員)

1992년 : 중앙대학교 컴퓨터공학과 공학사. 1994년 : 중앙대학교 대학원 컴퓨터공학과 공학석사. 2000년 1월 ~ 현재 : 중앙대학교 첨단영상대학원 연구교수 재직. <주관심분야 : 컴퓨터그래픽스, 멀티미디어 응용, 컴퓨터비전 등.>

洪顯基(正會員)

1993년 : 중앙대학교 전자공학과 공학사. 1995년~1998년 : 중앙대학교 대학원 전자공학과 공학석사, 공학박사. 1998년 9월~1999년 8월 : 서울대학교 자동제어특화연구센터 연구원. 1999년 9월~2000년 2월 : 중앙대학교 정보통신연구소 연구교수. 2002년 2월~2003년 1월 : Univ. of Colorado at Denver 방문 연구원. 2000년 3월~현재 : 중앙대학교 첨단영상대학원 영상공학과 조교수 재직. <주관심분야 : 컴퓨터 비전, 컴퓨터그래픽스, 멀티미디어 응용 등>