

증강현실을 위한 매트릭 복원

Metric Reconstruction for Augmented Reality

유정재, 김혜미, 박창준, 김홍석, 이인호
한국전자통신연구원 디지털콘텐츠 연구단

요약 이 논문에서는 영화, CF 같은 영상물 제작 시 CG/실사 합성을 위해 배경기하정보를 추출하는 알고리즘을 제안한다. Metric Reconstruction 은 카메라 자동 보정을 통해 이루어지며 이는 오랫동안 연구되어 온 분야이다. 접근방법은 영상의 특징점 추적 정보와 카메라 내부변수 가정으로부터 유도되는 자기 보정 방식과 공간상에서 미리 기하 정보를 알고 있는 보정들을 사용하는 방식으로 크게 분류될 수 있다. CG/실사 합성의 작업 효율성을 위해서는 배경 영상에 보정들이 보이지 않는 것이 좋은데 자연 특징점(Natural Feature)에만 의존하는 자기 보정 방식의 경우 2K 급 영상에서 CG 객체를 합성했을 때 떨림이 느껴지지 않을 만큼 정확한 결과를 얻기 힘들다. 이 논문에서는 Pollefeys[2]가 제안하였던 영상 시퀀스를 입력으로 하는 자기 보정 시스템을 바탕으로 마야 작업 환경에서의 핀홀 카메라 모델에 맞도록 카메라 내부변수의 비선형 최적화를 수행하는 방법과 사용자 개입을 통한 카메라 변수 정확도 향상방법을 제안한다.

핵심어 : Metric Reconstruction, 카메라 자동 보정, CG, 합성, 배경기하

1. 서론

본 논문에서는 2K급 영화 제작 시 CG/실사 합성에 사용될 수 있는 Metric Reconstruction 알고리즘을 제안한다. 카메라 자동보정과 Metric Reconstruction은 서로 맞물려 있는 관계이며 이 분야는 근래 20여년간 활발한 연구가 진행되어 왔다. [6]은 카메라 자동보정에 관한 전반적인 흐름을 설명하였고 [7]은 Pollefeys가 십여년간 영상 시퀀스 기반 카메라 자동보정에 관해 연구해 온 결과를 집대성한 논문으로 볼 수 있다. 이 논문은 주된 내용은 Hartley와 Pollefeys의 연구 결과들([1][4][7])을 참조하여 시퀀스 영상을 입력으로 하는 카메라 자동보정 시스템을 구현한 결과이며 성능 향상을 위해서 중간중간 약간씩의 새로운 방법들을 제안하였다. 또한 마야 작업 환경에서 컴퓨터 그래픽 객체와 실사영상을 합성하기 위해서 마야가 가정하는 핀홀 카메라 모델 형태로 카메라 내부, 외부 변수를 수정할 필요가 있다. 이 논문에서는 Local minimum에 빠지지 않도록 내부변수를 단계적으로 조정하는 방법을 제안하였다.

이 논문에서는 카메라 상태변수에 대한 가정을 제약조건으로 DAQ(Dual Absolute Quadric)의 일관성에 기반하여 Metric Reconstruction을 수행한다. 실험은 VX 2000으로 직접 촬영한 실내 영상과 부주3.0에서 제공하는 튜토리얼 영상에 대해서 수행하였다. ㄷ

2. 본론

2.1 Projective Reconstruction

본 연구에서 주된 관심은 기하복원이므로 특징점 추출 및 추적은 카메라 트래킹 상용 소프트웨어인 Boujou 3.0을 사용하였다. 특징점 데이터 추적 데이터가 입력되면 먼저 처음 두 개의 key frame 1,5 프레임 데이터에서 Normalized Eight Point Algorithm[1]을 사용해서 F를 계산하고 최적화 과정을 거친다. 이 과정에서 오정합된 특징점의 영향을 줄이기 위해서 RANSAC[3]을 사용하였고 최적화는 LM알고리즘을 사용한다. 이렇게 계산된 F를 초기값으로 F의 최적값을 예측하는 Gold standard algorithm[1]을 거쳐서 point correspondence에 관한 Geometric error를 최소화하는 F를 계산한다. 이렇게 계산된 F로부터 아래 수식 1과 같이 P1, P5를 선택한다.

$$P_1 = [I | \mathbf{0}]$$
$$P_5 = [\mathbf{e}_x F + \mathbf{e}\mathbf{v}^T, \mathbf{r}] \quad \text{수식 1}$$

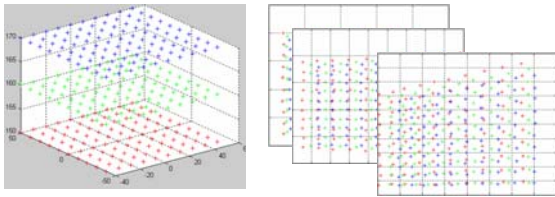
이때 \mathbf{e}_x 는 epipole \mathbf{e} 의 skew matrix 형태이며 외적 연산과 동일한 의미이다. \mathbf{v} 와 \mathbf{r} 은 rank를 늘이기 위해 첨가하는 임의의 3×1 벡터와 스칼라이다. 이렇게 구한 P1,P5로부터 Triangulation[4]을 사용해서 3D 상의 좌표 X를 구하며 {X}와 2,3,4 영상에서의 특징점 좌표를 이용하여 SVD를 거쳐 P2,P3,P4를 계산한다.

2.2 Bundle Adjustment

앞에서 복원된 X를 다시 역사영시켜 얻는 이미지들은 본래의 각 영상에서의 특징점 위치들과 오차를 갖게 되며 이를 RPE(Reprojection Error)라고 하고 Projective Reconstruction의 성능을 평가하는 지표로 삼는다. 무리조정(Bundle Adjustment)은 Sparse Matrix와 LM알고리즘[5]을 통해 수식 2에서 ε 를 최소화하는 σ 를 구하기 위해 iteration을 반복하는 과정이다. 수식 2에서 J=Jacobian matrix, σ =parameter set, ε =cost vector에 해당한다. 실제 실험은 매트랩상에서 최적화 함수와 sparse matrix 툴을 사용하여 수행하였다.

$$\begin{aligned}
 J\sigma &= \varepsilon \\
 \sigma &= [\mathbf{p}_2^T, \mathbf{p}_3^T, X_1^T, X_2^T, \dots, X_n^T]^T \\
 \varepsilon &= [(\mathbf{x}_1^1 - \mathbf{x}'_1)^T, \dots, (\mathbf{x}_n^1 - \mathbf{x}'_n)^T, \\
 &\quad (\mathbf{x}_1^2 - \mathbf{x}'_1)^T, \dots, (\mathbf{x}_n^2 - \mathbf{x}'_n)^T]^T \\
 X_i &= \text{ith feature point in 3D space} \\
 \mathbf{x}_i^j &= \text{ith feature point in jth image}
 \end{aligned}
 \tag{수식 2}$$

수식 2에서 p2,p3는 3*4 matrix형태의 카메라 사영행렬을 12*1 형태의 vector로 재정렬한 것이고 \mathbf{x} , X는 특징점의 이미지상의 좌표와 공간상의 좌표를 나타내는 3*1, 4*1 vector이다.



(a) 3D공간상의 분포 (b) P1, P2, P3 의 이미지

그림 1. Synthetic Data

그림 1은 같은 가상의 데이터 이미지에 노이즈를 가하고 2.1절의 방법으로 P1, P2, P3와 3D 복원된 X를 구한 결과이다. 실제 시스템의 작업 흐름에서는 Projective Reconstruction과 Metric Reconstruction후에 각각 번들조정이 수행된다.

2.3 Sequential Projective Metric Reconstruction

일반적으로 base line이 길어질수록 3D 복원 결과는 정확해지지만 2D상에서 feature tracking은 이어지는 특징점을 찾기 어렵게 된다. 반대로 base line이 짧을 경우에는 feature tracking은 쉬어지지만 3D 복원 결과는 노이즈에 취약하게 된다. 이러한 점을 해결하기 위하여 본 논문에서는 그다지 길지 않은 초기 구간에서 F를 계산하여 Projective Reconstruction을 수행하고 이후 프레임에서는 이전 구간에서 구한 Projective Reconstruction 결과에 기반해서 선형적인 방법으로 카메라 행렬을 예측하는 방법을 사용하였다. 이 과정에서 역시 최적화 작업이 첨가되며 각 프레임에서 카메

라 사영행렬을 구할 때 특징점을 inlier와 outlier로 분류하기 위하여 본 논문에서는 최소 중간 제곱법 (Least Median Square)을 사용하였다.

2.4 Metric Reconstruction

이 논문에서는 Hartley가 제안한 DAQ(Dual Absolute Quadric)[1]에 기반한 방법을 활용하였다. Metric Reconstruction은 영상에 보이는 대상물체의 3차원 공간 정보를 복원하는 과정이며 이는 카메라의 내부변수와 외부변수를 예측하는 카메라 자동보정과 서로 맞물리는 관계에 있다. 카메라 자동 보정을 개념적으로 설명하자면 다음과 같다. Projective Reconstruction결과에서는 Up to projective transform의 범위안에서 여러 형태의 복원 형태가 가능하며 Metric Reconstruction은 이들 중 일반적인 카메라 성질에 가장 적합한 형태를 선택하는 과정으로 이해할 수 있다.

일반적으로 내부변수 행렬 K는 각각의 프레임에서 변화 가능하다. 우리는 Absolute conic의 이미지가 모든 프레임에서 일관성 있게 보여야 한다는 성질을 이용하여 여러 프레임에서 Projective Reconstruction된 결과로부터 K를 예측할 수 있다.

이 과정은 수식 3에서 w^{*i} 의 각항에 대한 가정으로부터 DAQ를 선형적으로 구하는 방법이다. w^{*i} 은 i번째 프레임에서 Image of dual absolute conic이며 이는 카메라의 위치, 자세에 관계없이 내부변수 변화에만 의존한다. 자세한 설명은 [1]를 참조하기 바란다.

$$w^{*i} = K^i K^{iT} = P^i Q_{\infty}^* P^{iT}
 \tag{수식 3}$$

w^{*i} : Image of absolute conic of ith image

Q_{∞}^* : Dual absolute quadric of ith image

P^i : Camera projection matrix of ith image

K^i : Intrinsic parameter matrix of ith image

이 논문에서 추가로 제안하는 방법은 자동 보정 단계에서 사용자가 개입하여 공간상에서 서로 수직인 직선성분을 지정해 주고 수식 4에서 $\cos\theta=0$ 이 된다는 관계로부터 DAQ를 구하기 위한 constraint를 한 행 추가해 주는 것이다.

$$\cos \theta = \frac{\mathbf{v}_1^T w \mathbf{v}_2}{\sqrt{(\mathbf{v}_1^T w \mathbf{v}_1)(\mathbf{v}_2^T w \mathbf{v}_2)}}
 \tag{수식 4}$$

위의 수식에서 $\mathbf{v}_1, \mathbf{v}_2$ 는 해당 직선방향의 소실점을 나타내며 이는 그림 2와 같이 먼저 사용자가 평행한 직선의 ROI(Region of interest)를 지정하고 Radon Transform[8]을 사용해서 직선성분을 추출한 뒤 이들의 소실점을 예측하여 사용하였다. Radon Transform을 수행하는 과정은 알고리즘 1과 같다.

- Get the first image of the sequence.
- Define the rectangular "Region(s) Of Interest" for the whole image sequence by pointing upper-left and lower-right pixels to narrow down the search space.

- c. Convert ROIs into grayscale.
- d. Find edges by applying Canny Edge Operator into the grayscaled ROIs.
- e. Apply Radon Transform [2].

f. Detect and return the coordinate (θ, x') which maximizes the value

$$\Re[f](\theta, x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - x') dx dy$$

where x' is the shortest distance from the line to the origin and θ is the angle the line makes with the y -axis.

g. If the current frame is not the final, get the next image and go to step c.

알고리즘 1. Radon Transform을 이용한 직선성분 검출

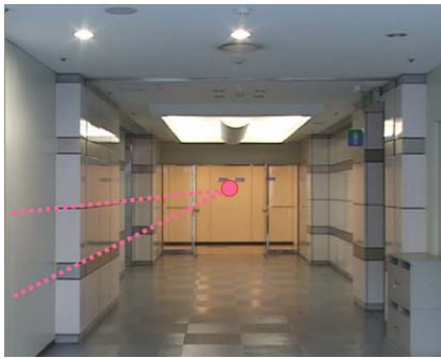


그림 2. 소실점 계산

2.5 Non-linear optimization of K

이 논문에서의 최종 목적은 카메라 정보를 추출해서 컴퓨터 그래픽 객체와 실사 영상을 합성하는 것에 있고 이러한 최종 합성은 일반적으로 마야 작업 환경에서 이루어진다. 그림 3은 필자의 이전 연구에서 바다 영상에서 카메라 움직임을 예측해서 CG 군함을 실사 바다 영상과 합성했던 작업 화면이며 그림 4는 그렇게 합성하여 렌더링 과정을 거친 결과이다.

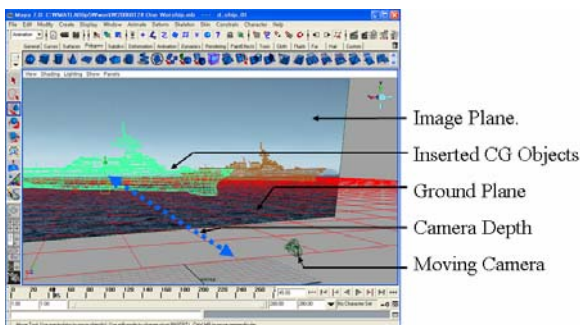


그림 3. 마야 작업환경에서 CG객체와 실사 영상의 합성



그림 4. CG 객체와 실사영상 합성 후 렌더링 결과

문제는 이와 같이 카메라 정보를 마야 작업환경으로 Importing하기 위해서는 마야에서 가정하고 있는 핀홀 카메라 모델과 일치하도록 카메라 내부변수 형태를 조절해야 한다는 점이다. 2.4절에서 카메라 내부변수에 관한 일반적인 제약조건에 기반해서 자동보정을 수행하였지만 이 과정에서 skew, aspect ratio, principal point 등에 관한 제약은 soft constraint이기 때문에 결과로 얻어지는 값은 핀홀 카메라 모델과 다르게 된다. 따라서 카메라 사영행렬을 카메라 내부변수와 외부변수 형태로 분할한 뒤, 아래 수식 5의 꼴로 나타내어지는 비용함수를 줄이는 방향으로 최적화를 수행해야 한다. 여기서 말하는 내부변수의 최적화란 결국 초점 거리를 제외한 내부변수들을 자동보정 단계에서의 soft constraint에 맞게 단순화 시키는 과정을 의미한다.

$$\text{cost function} = \sum_{i=1}^N \left(f(\mathbf{x}_i) * d(\mathbf{x}_i - \tilde{\mathbf{x}}_i) \right) + N * Wi * \left((r-1)^2 + 100s^2 + u^2 + v^2 \right)$$

수식 5

N 은 해당 프레임에서 영상안에 보이는 특징점의 개수이고 $f(\mathbf{x}_i)$ 는 \mathbf{x}_i 가 카메라 사영행렬을 구하는 과정에서 inlier인지 outlier인지 구분해주는 함수, N' 은 그렇게 선택된 inlier 특징점의 개수를 의미한다. r 은 aspect ratio, s 는 skew, (u, v) 는 principal point 좌표이다. Wi 는 내부변수 조건에 대한 가중치이며 초기에는 이 값을 작게 두어 단순히 역사영 오차를 최소화 하는 변들조정으로 시작하고 점차 이 가중치를 높임으로써 Local minimum에 빠지지 않으면서 결국 $r=1, s=u=v=0$ 형태의 카메라 사영행렬을 얻게 된다. 정리하여 말하자면 수식 5에서 첫번째 줄은 역사영오차에 관한 항이고 두번째 줄은 내부변수 최적화에 관한 항이다. 내부변수 조건 수식안의 계수는 [7]에서 제시한 내부변수 분산에 관한 사전지식을 참조하였다.

3. 실험결과

실험은 2단계로 진행되었다. 우선 영상 시퀀스에서 5개의 키프레임을 수동으로 선택하여 Projective Reconstruction, Bundle Adjustment, Metric Reconstruction을 순차적으로 수행하였다. 실험영상은 VX 2000 디지털 캠코더를 사용하여 실내에서 촬영한 영상을 사용하였고 실험결과 RPE (Reprojection Error)는 0.2 pixel 이었다. 소실점 정보를 활용하기 전, 후 RPE는 변함없지만 특징점의 공간 복원이 실제 공간 분포에 더 유사하게 되는 것을 확인하였다. 인텔 듀얼

코어 2.8 GHZ 환경에서 매트랩 6.5를 통해 실험하였고 특징점 추적을 제외한 연산시간은 12초였다. (특징점 추적은 부주 3.0을 사용) 그림5는 3개의 특징점들의 영상에서의 위치와 복원된 좌표간의 대응관계를 도시한 것이다.

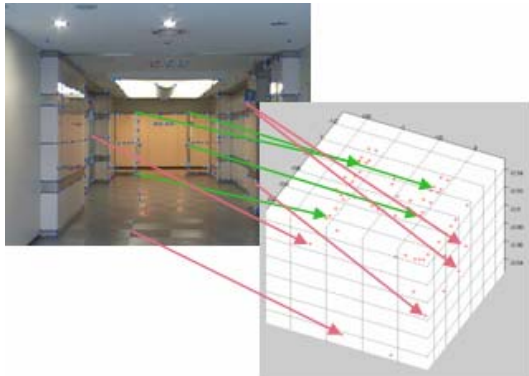


그림5. 특징점의 3D 좌표복원

두번째 단계에서는 전체 시퀀스 영상을 입력으로 초기 구간에서 Fundamental Matrix를 계산해서 Projective Reconstruction을 수행하고 이후 프레임에서는 앞의 결과에 맞도록 선형적인 방법과 최적화 방법을 병행하여 카메라 사영행렬을 계산하는 실험이었다. 실험 영상은 부주 3.0의 튜토리얼 1번 영상중 초기 30프레임을 사용하였다. 기하복원을 위한 특징점 추적데이터는 역시 부주 3.0에서 연산한 결과를 사용하였으며 그림 6은 부주 3.0의 작업환경을 보여준다. 매트랩 상에서 연산시간은 동일환경에서 한 시간 정도 소요되었으나 이는 C코드상에서 최적화 구현시 상당부분 감소시킬 수 있을 것으로 예상된다.

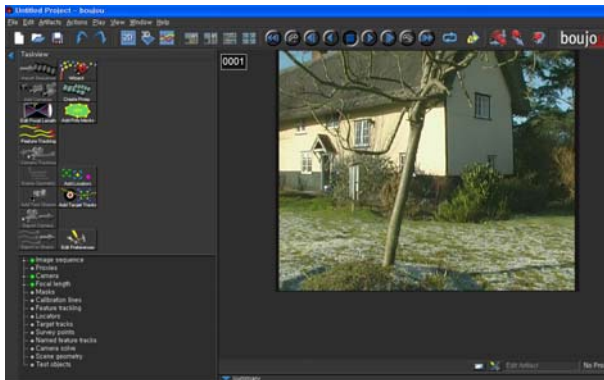


그림 6 부주3.0의 작업환경

그림 7은 최종 결과를 매트랩 상에서 3차원 도시한 결과이다. 보다시피 건물의 벽면과 땅 평면간의 직각 관계가 거의 정확하게 복원된 것을 확인 할 수 있다.

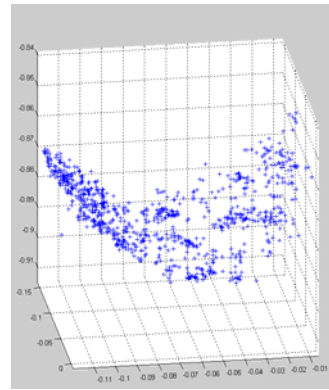


그림 7. 3차원 복원 결과

일반적인 CG/실사 합성에서 소스 영상만을 입력으로 정확한 카메라 내부, 외부 변수와 기하정보를 자동으로 추출한다는 것은 매우 어려운 일이다. 영화제작 현장에서 이슈는 완전히 자동화할 수 있는가 보다도 얼마만큼 떨림 없이 정확한 결과를 얻을 수 있는가의 문제이다. 이번 연구의 의의는 기존의 연구내용들을 참조하여 DAQ의 일관성에 기반하는 자동 보정방식을 구현하면서 마야에 Importing 하여 CG/실사 영상 합성을 수행할 수 있는 핀홀 카메라 모델의 예측에 있다고 볼 수 있다. 앞으로 남은 일은 2K급 고화질 영상에서 카메라 모션이 격렬한 경우에도 CG 객체와 실사 배경간의 떨림을 감지할 수 없을 만큼 정확한 카메라 정보를 추출하는 방법을 개발하는 것이다.

참고문헌

- [1] "Multiple View Geometry in Computer Vision," Richard Hartley
- [2] "Tutorial on 3D Modeling from Images," Marc Pollefeys, <http://www.esat.kuleuven.ac.be/~pollefeys/tutorial/>
- [3] "An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration," A.J. Lacey, N. Pinitkarn and N.A. Thacker, *BMVC 2000*
- [4] "Triangulation," Richard I. Hartley, Peter Sturm
- [5] "The Levenberg-Marquardt Algorithm," Ananth Ranganathan
- [6] "A Survey of Camera Self-Calibration," Elsayed E. Hemayed, *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2003)*
- [7] "Visual Modeling with a Hand-Held Camera," Marc Pollefeys, *IJCV 59(3), p207-232, 2004*
- [8] "The Radon Transform and Some of Its Applications," DEANS, S. R., *John Wiley & Sons, 1983.*