

증강 비디오 시스템을 위한 AR 영상 Registration 알고리즘 연구

김혜경^o, 오해석

승실대학교 컴퓨터학과 멀티미디어연구소

hkkim@multi.soongsil.ac.kr, oh@computing.soongsil.ac.kr

A Study of AR Image Registration Algorithm For Augmentation Video System

Hye-Kyung Kim^o, Hae-Seok Oh

Multimedia Lab., Dept. of Computer Science, Soongsil University

요 약

본 논문에서는 비디오 영상열 내에 컴퓨터가 생성한 가상의 3D 영상을 이음새 없이 추가하기 위한 문제에 초점을 맞추고 있다. 2단계의 견고한 통계적인 메소드는 추적된 커브들의 모델-영상 대응점으로부터 보다 정확한 자세를 평가하기 위하여 자세 계산을 위해 사용되었다. 또한, 관점의 정확성 향상을 위하여 두 개의 연속하는 영상들간에 매치될 수 있는 핵심점을 카메라 움직임에 대한 상관관계 함수로 사용하여 매칭 에러와 reprojection 에러를 포함한 비용함수를 최소화함에 의해 관점을 향상시킨다. 비디오 영상내 객체 영상과 가상의 3D 영상간에 발생하는 폐색 공간 문제를 해결하기 위하여 반 자동 알고리즘을 제안하였다.

1. 서론

증강현실(AR)은 컴퓨터 기반의 정보와 데이터베이스의 가능성을 이용하고 개발하기 위한 매우 효율적인 수단이다. 그러므로, 효과적인 AR 비디오 시스템을 만들기 위하여, 컴퓨터가 생성한 객체들과 실제 비디오 영상 장면은 반드시 이음새 없이 추가되어야 한다. 또한 비디오 영상의 효과를 증강시키기 위하여 영상의 떨림 현상과 노이즈를 최소화하는 문제는 필수적인 문제이다. 인간의 시각 시스템은 매우 작은 mis-registration도 감지하므로 registration 작업은 특별한 주의를 필요로 한다. 가상객체와 실제 객체들의 현실적인 병합은 객체들이 주위환경에서 물리적으로 그럴듯한 방식으로 행동하기를 요구한다. 즉, 비디오 장면내 객체에 의해 가려질 수 있고, 다른 객체에 의해 그림자가 지는 현상이 표현되어야 한다.

2. AR 영상 Registration 시스템의 전체적인 개요

증강 비디오 영상을 위하여 제안된 Registration 시스템은 전체적으로 4단계로 구성한다. 즉,

- 1) 트래킹평가와 자세 계산 메소드를 사용한 관점 계산
- 2) 2D/2D 대응점 계산
- 3) Mixing한 가상의 객체 영상 추가
- 4) 실제 비디오 영상내에 컴퓨터가 생성

2.1 포인트 대응점으로부터 자세 계산

카메라의 정확한 자세를 계산하기 위해 rotation R 과 translation t를 발견하여 비디오 영상내의 세계 좌표 시스템을 카메라 좌표 시스템으로 나타낸다. 그러므로, 카메라의 내부 파라미터들을 캘리브레이션 프로세스에 의해 측정하기 위해, 6개의 파라미터들을 결정해야 하

고(R에 대해 3개, T에 대해 3개), 벡터 p 에 의해 표현했다. 여기서 3D 포인트들 $M_{i(1 \leq i \leq n)}$ 를 알고 있다고 가정하고 이미지 내에서 그들의 대응하는 키포인트들을 $m_{i(1 \leq i \leq n)}$ 라고 가정한다. 관점을 계산하는 것은 R, T를 찾는 것과 같으며 그것은 re-projection 에러를 최소화한다. 즉,

$$\min_{R, T} \sum r_i^2 = \min_{R, T} \sum dist(proj(M_i), m_i)^2$$

최소의 추정량은 가짜 일치에 강건하지 않으므로, 가장 강건한 기법중의 하나인 M-estimator를 사용한다. 즉, 제공된 나머지를 대신하여 최소화하는 유도함수를 식 (1)과 같이 사용한다.

$$\min_p \sum_{i=1}^n \rho(r_i) \quad (1)$$

(1)을 최소화하는 것은 p 에 대한 초기 평가를 사용한 표준 기법을 수행해야 하므로, 본 논문에서는 Powell's 메소드[1]를 사용한다.

3D 영상의 커브들에 대응하는 2D 영상의 커브들간의 point-to-point의 대응점들은 이용할 수 없기 때문에 관점 계산은 더욱 더 어려운 문제이다. 그러므로, 견고한 추정을 위해 2단계 통계적인 메소드를 제안하였다. 즉, local 단계는 각 특성에 대한 나머지를 계산하고, global 단계는 나머지 함수를 최소화한다.

• local 단계 : 이 단계의 목표는 틀린 부분에 대한 영향을 줄이는 것이다. 즉, 영상 커브 C_i 의 나머지 에러 r_i 는 거리 $\{d_{i,j}\}_{1 \leq j \leq l}$ 에 대한 견고한 함수에 의해 계산된다. 즉,

$$r_i^2 = \frac{1}{l_i} \sum_{j=1}^l \rho(d_{i,j}). \quad (2)$$

• global 단계 : 견고한 나머지가 각 특성에 대해 계산 되었을 때, 관점은 최소화하여 계산한다.

$$\min_p \sum_{i=1}^n \rho(r_i) \quad (3)$$

그러므로, 완벽하게 틀린 특성들은 나머지가 커지게 되므로 버린다.

2.2 관점 계산

두 개의 일치된 포인트들간의 관계는 카메라 움직임 함수와 내부 파라미터 함수이므로, 관점은 두 프레임들간의 매칭 에러뿐만 아니라 reprojection 에러를 둘러싼 cost 함수를 최소화함에 의해 향상시킬 수 있다. 키 포인트들이 일반적으로 모델 포인트들에 대응하지 않은 이래, 관점 계산은 포인트들의 2D 대응을 통해 향상시킬 수 있다. 그러므로, 특징점을 추출하는 measure를 계산하기 위해 Harris and Stephens[2]의 자동상관함수를 사용한다. 즉, 행렬의 고유치는 자동상관함수의 주요한 곡률로, 이 값이 높으면 키포인트가 추출된다.

$$\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \left(I_x = \frac{\partial I}{\partial x} \dots \right) \quad (4)$$

또한, 연속하는 두 영상간의 키 포인트 매칭을 위해 사용하는 방법은 아래와 같은 차별적인 불변량 벡터를 사용하여 중요한 기하적 변환에서도 매칭을 더욱 쉽게 만든다.

$$\begin{bmatrix} I \\ I_x^2 + I_y^2 \\ I_{xx} I_x^2 + 2 I_{xy} I_x I_y + I_{yy} I_y^2 \\ I_{xx} + I_{yy} \\ I_{xx}^2 + 2 I_{xy} I_{yx} + I_{yy}^2 \end{bmatrix} \quad (5)$$

2.3 Mixing 메소드

비용함수를 최소화하기 전에, 두 개의 일치된 키-포인트 q_1, q_2 와 두 개의 viewpoint $[R_k, t_k]$ 과 $[R_{k+1}, t_{k+1}]$ 사이의 관계 항목을 불러들인다.

$\Delta R, \Delta t$ 를 k 프레임과 k+1 프레임간의 카메라에 대한 상대적인 이동이라고 하고, A는 카메라의 내부 파라미터 행렬이라고 하자. 즉,

$$A = \begin{bmatrix} k_u f & 0 & u_0 \\ 0 & k_v f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

흥미있는 포인트를 사용하여 관점계산을 향상하기 위한 간단한 방법은 다음을 최소화한다.

$$\min_{R_{k+1}, t_{k+1}} \left(\frac{1}{n} \sum_{i=1}^n \rho_1(r_i) + \frac{\lambda}{m} \sum_{i=1}^m \rho_2(v_i) \right) \quad (7)$$

• r_i 는 프레임 k+1 내에서 추적된 특성들과 모델의 프로젝트된 특성들간의 거리이다.

• v_i 는 q_1^i 과 q_2^i 간의 매칭 품질을 측정한다.

• ρ_1 과 ρ_2 는 키포인트 대응을 위하여 사용된 M-추정량이다.

• λ 파라미터는 이용 가능한 3D 데이터에 속하는 근사와 키-포인트들간의 2D 대응점들의 품질 사이에서 타협안을 제어한다. 우리는 실험에서 $\lambda=1$ 을 사용하였다.

등식3의 최소화는 Powell's 알고리즘과 같이 최소화를 위해 반복적인 알고리즘을 사용함에 의해 계산된다.

3. Occlusion 문제 해결

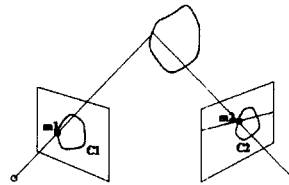
가상의 영상과 실제 영상의 현실감있는 결합은 실제 환경안에서 물리적으로 그럴듯한 방식으로 행동하는 것을 요구한다. 즉, 장면내의 객체에 의해 가상의 영상이 가려지거나 그림자로 덮혀지는 Occlusion 문제를 해결하기 위해서는 실제 장면의 깊이에 대하여 가상 객체의 깊이를 비교하는 작업이 필요하다. 그러나, 회복된 맵의 정확성이나 밀도가 occlusion 경계선 추정에 충분하지 않고, 또한 두 프레임간의 카메라 움직임은 실제 AR 애플리케이션에서 완벽하게 알지 못하므로 매우 어려운 문제이다. 그러나 반드시 해결이 되어야 할 문제이다. 그러므로, 본 논문에서 폐색공간 문제를 해결하기 위해 제안된 메소드는 다음과 같다.

1)사용자가 선택된 프레임들간의 집합에서 감춰진 윤곽을 그린다. 2)2개의 연속하는 키 프레임으로부터 감추어진 객체의 3D의 폐색 경계선을 구축한다. 3)3D 곡선의 프로젝트션을 사용해서 2개의 핵심 뷰들간의 프레임에서 2D 감추어진 경계선을 예측한다. 4)예측된 경계선은 정제되고, 회복된 경계 포인트들이 그들 영역에 반드시 존재해야하고, 영역기반 트래킹과 active contour model을 사용한다.

•관점의 불명확성 추정 : 관점 P 는 함수 $\Phi(P)$ 를 최소화함에 의해 회복되었다.

$$\Phi(p) \approx \Phi(p^*) + \nabla \Phi(p^*)^t \delta p + \frac{1}{2} \delta p^t H(p^*) \delta p \quad (8)$$

•3D 감추어진 경계선 구축 : occluding 객체의 경계선 C1과 C2는 두 개의 연속하는 중심 뷰에서 윤곽선을 그렸고, 3D occluding 경계를 3각의 stereo 측량을 사용하여 구축한다. 즉, m1은 c1 위의 포인트라 가정하면, 다른 뷰에서 그것의 대응하는 포인트 m2는 C2 와 함께 epipolar line 에 대한 교선이 된다. 일반적으로, 순위 제약은 대응 문제의 애매성을 해결하기 위해 사용되었다.



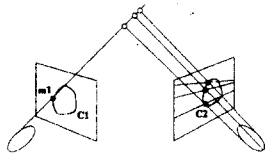
•관점위의 불명확성 제거

관점위의 불명확성을 고려하기 위하여 두 개의 중심

뷰에서 외부의 관점을 고찰한다. 즉, 외부의 관점은 다음과 같다.

$$\{p_1^1, \dots, p_1^{12}\} \text{ (resp } \{p_2^1, \dots, p_2^{12}\}) \quad (9)$$

C1 위의 포인트를 m_1 이라 하면, 주어진 외부 관점을 p_1 , 중심 뷰 2에서 12 개의 극단적인 뷰들과 함께 m_1 에 대하여 12개의 가능한 복원을 계산할 수 있다.



윤곽선 재인 알고리즘

컴퓨터가 생성한 객체 V 에 해당하는 영역에서 추출된 윤곽선들은 다음 영상내에서 자취를 남긴다. 두 영상들 간의 윤곽선 포인트들의 매칭은 epipolar 제약을 사용하여 실행되고, 두 대응점들은 결국 (m_1, m_2) 로 표현된다. 각 contour point m_1 을 장면과 컴퓨터가 생성한 객체에 대한 대응점의 상대적인 위치에 의존하여 '앞' 이나 '뒤' 로 이름이 부여되는 labeling 단계를 거친다. 이런 목적을 위하여, f_{m_1} 을 다음처럼 정의한다.

$$f_{m_1}: Z \rightarrow proj_{I_2}(m_{1x}, m_{1y}, Z)$$

여기서 $m_1 = (m_{1x}, m_{1y})$ 이고, $proj_{I_2}$ 는 영상 I_2 내의 프로젝션이다. $f_{m_1}(Z)$ 는 calibration 프로세스와 영상 포인트 m_1 에 의존하는 유사 함수 Z 의 계수로써 표현될 수 있다.

$$m_2 = f_{m_1}(Z_{real}) \text{ and } m_{obj} = f_{m_1}(Z_{obj})$$

occluding 객체들 식별

근접 조건을 기반으로 한 두 개의 커브 C_i 와 C_j 에 대한 거리는 스톨레스홀드 s 가 주어지고, 두 개의 체인이 $distance(C_i, C_j) < s$ 일 때 동일한 객체에 속한다고 정의한다.

근접 그래프 g 를 만든다. : 즉, 체인 $C_i (1 \leq i \leq N)$ 로 표현한 노드들은 두 개의 노드에 대응하는 체인들간의 거리가 s 보다 적을 때만 연결된다. 그러므로, occluding 객체들을 감지하는 것은 근접 그래프내의 소집단을 계산하는 것과 같다. 그것은 커브의 집합 $HC \subset g$ 이다. 즉, $\forall C \in H, \exists C' \in H \mid C \text{ 와 } C' \text{ 연결됨.}$

occlusion mask 계산

윤곽선의 집합 H 로부터 occlusion mask를 추론하는 작업은 영역 기반 트래킹과 함께 active contour model을 사용한다. 그러므로 최소화 될 함수가 다음과 같이 정의된다.

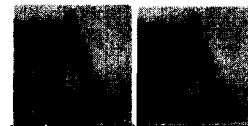
$$\phi_a(i) = \begin{cases} \sum_d (I_{template}(m_i+d) - I(trans f_a(m_i+d)))^2 & \text{if } trans f_a(m_i) \in A_i, \\ \alpha W^2 & \text{otherwise.} \end{cases}$$

$I_{template}$ 은 예견된 템플릿을 나타내고, I 는 현재의 이미지를 표현한다. α 는 고정 상수를 사용한다.

또한, occluding 경계의 조정을 위해 사용될 contour 모델은 다음과 같이 정의된다.

$$trans f_a(C) \quad (10)$$

4. 실험 결과



[그림 1] 회복된 occluding 경계



[그림 2] 가상객체의 occlusion 이 적용된 AR 장면

5. 결론 및 향후 연구 과제

본 논문에서는 AR 비디오 시스템을 위한 occlusion 문제를 해결하기 위한 새로운 방법을 제안하였으며, 실험으로 가상의 객체 영상을 증강시켜 입증하였다. 가장 큰 장점을 두 프레임사이에서 계산된 움직임 위의 불명확성을 제거하는 방법에 집중한다. 이것은 occluding 경계가 검색될 수 있는 흥미있는 영역을 정의하도록 허락하고 영역 기반 트래킹 알고리즘에 제약을 가한다. 키 프레임의 신중한 선택을 통해서, 다른 접근 방법보다 더욱 쉽고 정확한 AR 영상을 얻을 수 있다. 향후에는 좀더 다양한 영상을 대상으로 metric 정보를 추가한 증강된 비디오 영상을 제작하고자 한다.

6. 참고 문헌

[1] M. Berger, "Compositing Computer and Video Image Sequences," In Computer Graphics Forum, Conference Issue Eurographics', France, pp.23-32, August 1996.
 [2] D.Wang and C.Labit, "Morphological spatio-temporal simplification for video image segmentation," to appear in signal Processing: Image Communication.