

내부와 외부 카메라를 이용한 비전 기반 움직임 추정

*엄태영 **전지인 ***박종일
한양대학교

*uty02@mr.hanyang.ac.kr, **starlove7@mr.hanyang.ac.kr, ***jipark@hanyang.ac.kr

Vision Based Motion Estimation Method using Ego-Exo Cameras

*Uhm, Taeyoung **Jun, Ji-in ***Park, Jong-Il
Hanyang University

요약

최근, 인간과 컴퓨터 간의 상호작용을 위해 카메라의 정확한 포즈를 추정하고자 하는 연구가 많이 이루어지고 있다. 이러한 연구들은 인간의 움직임을 추적하기 위하여 카메라 영상으로부터 인간의 포즈를 추정하여 주된 인터랙션으로 활용하고자 한다. 그러나 기존의 움직임 추정 방법은 주로 내부(ego) 혹은 외부(exo)의 단일 카메라만을 이용하기 때문에 미세한 움직임을 분석하기 어렵다. 본 논문에서는 외부 카메라뿐만 아니라 내부 카메라를 혼합하여 사용함으로써 미세한 움직임도 추정할 수 있는 하이브리드 비전 기반 움직임 추정 방법을 제안한다. 실험 결과는 단일 카메라만을 이용한 결과와 비교해 더 정확한 포즈 추정을 보인다.

1. 서론

영상에서 모션을 추적하기 위해 카메라 포즈를 계산하는 방법은 많은 연구자들에 의해서 연구 되고 있다. 특히, 단일 혹은 멀티 카메라에서 마커를 인식하여 카메라 포즈를 계산하는 방법은 인간 컴퓨터 인터랙션 분야에서 크게 주목 받고 있다. 이러한 방법들은 대부분 외부 카메라를 이용하여 캘리브레이션을 수행하고 마커를 인식하는 방식을 주로 사용한다. 이러한 고정된 외부 카메라로부터 휴먼 포즈를 추정하는 방법은 마커와 거리가 멀거나 혹은 마커의 크기가 작은 경우에는 정확한 방향(오리엔테이션) 정보를 구하기 어렵다. 또한 로봇 메카트로닉스 분야에서의와 같이 외부 카메라의 반대 개념으로 로봇의 시점을 위해 내부 카메라를 사용하여 로봇의 움직임을 추정하는 연구들이 있다. 이러한 방법들은 주로 움직이는 카메라 영상에서 고정된 다수의 마커들을 이용하여 현재의 로봇 위치를 추정하는 방법을 사용한다. 하지만 내부 카메라 영상은 위치 추정에 있어서 근본적인 모호함을 갖고 있다. 이러한 내부와 외부 카메라를 이용한 각각의 연구는 계속 이루어져 왔다. 본 논문에서는 이러한 외부와 내부 카메라에 대한 비전 기술의 장점들을 함께 고려하여 단일카메라만으로 추정하기 힘든 미세한 움직임을 예측할 수 있고 비선형 방정식을 사용하지 않아도 보다 정확한 움직임 추정이 가능하도록 카메라의 영상을 이용하는 하이브리드 방법을 제안한다. 이는 외부 카메라에서 더욱 신용도 높은 translation 값을 예측하고 이를 기반으로 내부 카메라에서는 rotation 값을 예측하여 위치 추정 값을 더 정확한 값으로 보정한다. 따라서 미세한 움직임도 예측이 필요한 어플리케이션(ex: 재활치료를 위한 기능성 운동치료 게임<그림 1 참조>)에서도 사용자의 움직임을 추정하여 제공하는 것이 가능하다.

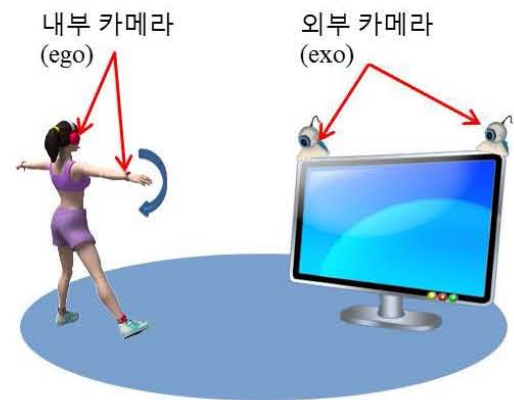


그림 1. 미세한 움직임 추정이 필요한 예시: 손목 운동을 포함한 운동 치료 게임.

2. 관련 연구

카메라를 이용한 모션을 예측하는 방법은 내부 카메라를 이용한 방법과 외부 카메라를 이용한 방법으로 크게 두 가지로 나눌 수 있다. 먼저 내부 카메라를 이용하는 방법은 주로 로봇이나 자동차의 움직임을 계산하기 위해서 사용한다. 이를 위한 방법 중에는 optical flow와 ToF 카메라를 이용하여 모션 벡터를 추출하는 방법이 있다[1]. 또한, 움직임을 예측해서 점근선 특성을 이용해 현재 카메라의 위치를 계산하는 방법도 있다[2]. 다음으로 외부 카메라를 이용하여 물체의 움직임에 대한 외부 파라미터를 계산하는 방법들도 optical flow를 이용한 방법[3]이 있다. 이러한 내부 혹은 외부 카메라를 이용한 대부분의 연구들은 보다 정확한 움직임 추정을 위해서 에리 함수를 모델링 하여 값

을 예측하는 방법들[4-6]을 오래전부터 연구 하였다. 하지만 외부 카메라 방법의 경우는 대상의 크기와 거리에 따라 방향성에 대한 모호함이 증가하고, 내부 카메라 방법의 경우는 근본적으로 모호함을 갖고 있다. 따라서 모션 정보를 알 수 있으나 미세한 움직임에서는 정확한 방향성(오리엔테이션)과 이동성(포지션) 값을 계산하기 어렵다. 또한 비선형 방정식을 적용하여 에러 값을 최소화 하여 추정하여야만 한다. 이와 같이 내부와 외부 카메라를 기반으로 움직임에 대한 외부 파라미터를 구하는 연구는 꾸준히 수행되어왔으나, 두 가지 카메라를 동시에 고려하는 연구는 아직 미비하다. 따라서 각각의 카메라에 대한 장점을 살려 미세한 움직임 추정뿐만 아니라 비선형 방정식의 적용이 필요 없는 정확한 움직임을 추정하고자 한다. 본 논문에서는 미세한 움직임이 필요한 인간 컴퓨터 상호작용을 목적으로 내부 카메라를 인간 움직임을 추정하기 위해 신체에 부착된 경우를 가정하고 있다.

3. 포즈의 중의성과 기하학적 해석

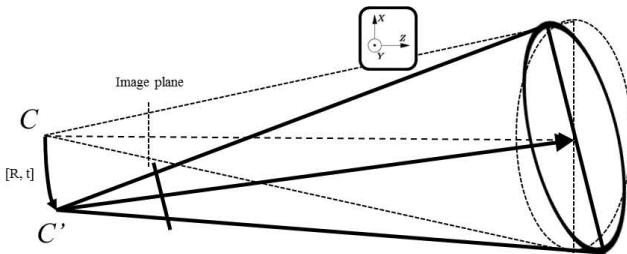


그림 2. 카메라 움직임에 의한 오리엔테이션과 포지션 모델.

미세한 모션을 예측하기 위해서는 정확한 카메라 포즈를 알고 있어야만 한다. 이는 캘리브레이션에 기반하여 내부 카메라의 6개로 이루어진 외부 파라미터 $R = [R_1, R_2, R_3]^T, t = [t_x, t_y, t_z]^T$ 를 구하는 문제에서 출발한다. 먼저 내부 카메라의 오리엔테이션 R 과 포지션 t 를 정의 하면 그림 2와 같이 투영된 카메라 센터와 이미지 평면에서 좌표계 모델을 이용해서 구할 수 있다. 이는 이미지의 센터에 위치한 카메라 모델의 중심점과 실제 카메라 센터가 같은 선상에 있다는 것을 가정하여 카메라 포지션 t 를 둘 사이의 거리 벡터로 정의할 수 있고 오리엔테이션은 카메라 모델의 rotation R 로 볼 수 있다. 이를 수식으로 정의 하면 내부 카메라 이미지의 픽셀 포인트 $p_{egi} = [p_{ix}, p_{iy}, 0]^T$ 에 대하여 내부 카메라 좌표계 v_{egi} 는 다음과 같다.

$$v_{egi} \propto R_{eg} p_{egi} + t_{eg} \quad (1)$$

이러한 내부 카메라의 포즈를 구하기 위해 마커 대부분의 방법과 같이 마커 기반 pose estimation 알고리즘을 이용하였다.

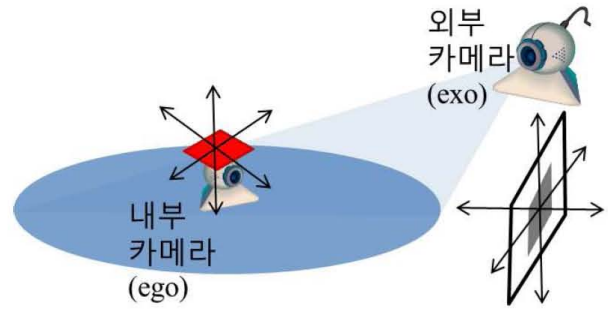


그림 3. 내부와 외부 카메라의 관계.

한편 내부 카메라 포즈와 역의 개념으로부터 외부 카메라 영상에서 내부카메라에 위치한 마커를 추출하여 내부 카메라의 포즈를 구할 수 있다. <그림 3 참조> 이는 외부 카메라 영상 픽셀 $p_{exi} = [p_{ix}, p_{iy}, 0]^T$ 과 좌표계 v_{exi} 로부터 마커의 움직임과의 관계로 추출한다.

$$v_{exi} \propto R_{ex} p_{exi} + t_{ex} \quad (2)$$

이를 통해 구한 마커의 포즈는 내부 카메라의 카메라 포즈와 대응된다. 하지만 각각의 카메라에서 구한 포즈 값은 외부 카메라에서의 포지션 t_{ex} 이 더욱 신뢰도가 있어 보인다. <그림5 참조> 따라서 가장 신뢰도가 높은 외부 카메라의 포지션 값을 기준으로 R_{eg} 값을 보정한다. 이는 내부 카메라의 움직임이 미세할 경우 카메라 포즈의 ambiguity의 error를 optimize하기 위해서 매우 유용한 방법 이다.

4. Robust pose estimation method

우리의 새로운 pose estimation 방법은 내부와 외부 카메라를 동시에 고려하는 것으로부터 시작된다.

내부 카메라의 파라미터들은 외부 카메라에 의해 수정된 카메라 포지션 정보를 기반으로 Projective transformation 행렬을 다음과 같이 정의 할 수 있다.

$$P = KG = K[R_{eg}, t_{eg}] \rightarrow K[R', t_{ex}] \quad (3)$$

따라서 새로운 t_{ex} 값에 대한 Rotation 파라미터를 구하기 위해 Homography를 이용한 다음과 같은 식을 사용한다.

$$K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} R'_{11} & R'_{12} & t_x \\ R'_{21} & R'_{22} & t_y \\ R'_{31} & R'_{32} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (4)$$

이는 평면으로부터 카메라의 포즈를 추정하기 위한 식

$$H \propto A[R, t] \quad (5)$$

에 기반하여 행렬 G 는 다음과 같이 정의 된다.

$$G = \lambda A^{-1}H \quad (6)$$

즉, 외부 카메라에 의한 포지션 t_{ex} 에 의해

$$\lambda' = [A^{-1}H]^{-1} \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \quad (7)$$

으로 λ' 값을 새로이 optimization 하여 G' 값으로 다음과 같이 보정 할 수 있다.

$$G' = \lambda' A^{-1}H \quad (8)$$

이렇게 구한 $G' = [G'_1, G'_2, G'_3]$ 은 다음과 같이 정규화를 한다.

$$G' = \left[\frac{G'_1}{l}, \frac{G'_2}{l}, \frac{G'_3}{l} \right] = [R'_1, R'_2, t_{ex}] \quad (9)$$

$$(l = \sqrt{\|G'_1\| \|G'_2\|})$$

보다 정밀하게 하기 위해, 아래의 식을 이용해서 Rotation 파라미터를 확정한다.

$$R_1 = \frac{1}{\sqrt{2}} \left(\frac{c}{\|c\|} + \frac{d}{\|d\|} \right), \quad (10)$$

$$R_2 = \frac{1}{\sqrt{2}} \left(\frac{c}{\|c\|} - \frac{d}{\|d\|} \right),$$

$$R_3 = R_1 \times R_2.$$

$$(c = R_1 + R_2, p = R_1 \times R_2, d = c \times p)$$

5. 실험 결과



그림 4. 실험 환경. (a) 외부카메라, (b) 내부 카메라.

실험은 <그림 4>와 같이 인간 움직임을 추정하기 위한 경우를 가정하여 내부 카메라는 320*240의 화질을 가진 웹캠을 이용하여 약한 캘리브레이션을 수행하였고, 외부 카메라는 HD급 화질의 웹캠을 이용하여 강한 캘리브레이션을 수행하였다. 두 카메라 간의 거리는 0.8m이

고 카메라와 마커 간의 거리는 0.85m로 고정하였다. 이에 따라 카메라의 포지션 정확도는 아래의 <그림 5>와 같다.

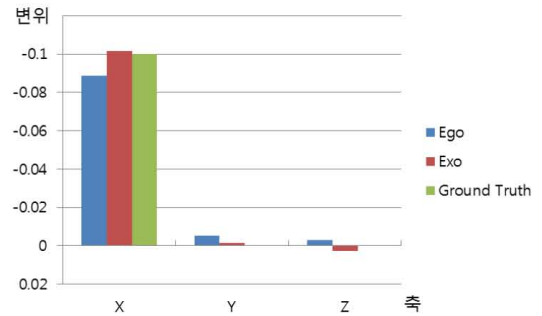


그림 5. 내부와 외부 카메라의 포지션 값의 정확도 비교. (X 축 방향으로 -0.1 만큼 이동한 경우)

따라서 외부 카메라의 포지션 정보를 기반으로 외부카메라의 오리엔테이션 값을 구하면 <그림 6>에서 보는 것과 같이 마커에 크기와 거리에 따라 오리엔테이션 오차의 성향을 보인다. 또한 내부 카메라만을 이용한 경우와 비교하여 내부와 외부의 하이브리드한 경우 오차는 0.2 미만으로 감소하는 것을 볼 수 있다.

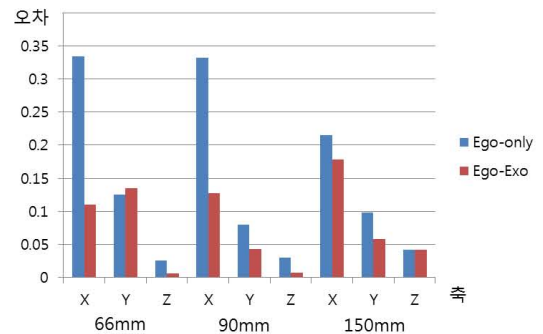


그림 6. 내부 카메라만을 이용한 경우와 하이브리드한 방법을 적용한 경우에 따른 오리엔테이션 오차 비교. (회전 없이 X축 방향으로 0.1 만큼 이동한 경우)

6. 결론 및 향후 과제

본 논문에서는 인간 움직임에 대한 추정을 위해 내부와 외부 카메라를 동시에 사용하여 외부 파라미터(포지션과 오리엔테이션)를 복잡한 비선형 에러모델 적용 없이 정확하게 구하는 방법을 제안하였다. 이는 내부와 외부 카메라 각각의 장점을 하이브리드 방법으로 적용하여 외부 카메라의 신뢰도 있는 포지션을 기반으로 내부 카메라의 오리엔테이션 정보를 더욱 정확하게 할 수 있도록 수정 할 수 있었다.

따라서 미세한 움직임을 포함하는 인간과 컴퓨터 인터랙션을 위해서는 인간 내부에 존재하는 카메라 영상과 인간을 바라보는 외부 카메라 영상을 하이브리드 방법으로 이용하면 더 정확한 움직임 추정이 가능하다. 향후에는 비 마커 기반 움직임 추정을 위해 외부 카메라는 스테레오 방법으로 내부 카메라는 optical flow 방법을 적용하고자 한다. 또한 Cramer-Rao Lower Bound와 같은 오차 분석 방법을 적용하

여 더욱 정확한 오차 비교를 수행할 필요가 있다.

참고문헌

- [1] S. Wang, H. Yu, and R. Hu "3D Video Based Segmentation and Motion Estimation with Active Surface Evolution," Journal of Signal Processing System, 2012.
- [2] J. Weng, Y. Cui, and N. Ahuja "Transitory Image Sequences, Asymptotic Properties, and Estimation of Motion and Structure," IEEE Transactions on Image Analysis and Machine Intelligence vol. 19(5), pp. 451-464, 1997.
- [3] C. Zach, T. Pock, and H. Bischof, "A Duality Based Approach for Realtime TV-L¹ Optical Flow," Proceeding of DAGM Symposium on Pattern Recognition, pp. 214-223, 2007.
- [4] J. Weng, T. S. Huang, and N. Ahuja "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," IEEE Transaction on Pattern Analysis and Machine Intelligence, pp.451-476, 1989.
- [5] J. Weng, N. Ahuja and T. S. Huang "Optimal Motion and Structure Estimation," IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 144-152, 1989.
- [6] S. Gerald and P. Axel, "Robust Pose Estimation from a Planar Target," IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 2024-2030, 2006.