가상 현실 어플리케이션을 위한 관성과 시각기반 하이브리드 트래킹

Hybrid Inertial and Vision-Based Tracking for VR applications

구 재필[1,2], 안 상철[2], 김 형곤[2], 김 익재[2], 구 열회[3]

[1] 연세 대학교 전기전자공학과(E-mail : kjpkoo@kist.re.kr)

[2] 한국과학기술연구원 영상미디어연구센터(E-mail : asc@cherry.kist.re.kr, hgk@kist.re.kr, kij@imrc.kist.re.kr)

[3] 서울대학교 전기공학과(E-mail : lm9102@hotmail.com)

In this paper, we present a hybrid inertial and vision-based tracking system for VR applications. One of the most important aspects of VR(Virtual Reality) is providing a correspondence between the physical and virtual world. As a result, accurate and real-time tracking of an object's position and orientation is a prerequisite for many applications in the Virtual Environments. Pure vision-based tracking has low jitter and high accuracy but cannot guarantee real-time pose recovery under all circumstances. Pure inertial tracking has high update rates and full 6DOF recovery but lacks long-term stability due to sensor noise. In order to overcome the individual drawbacks and to build better tracking system, we introduce the fusion of vision-based and inertial tracking. Sensor fusion makes the proposal tracking system robust, fast, accurate, and low jitter and noise. Hybrid tracking is implemented with Kalman Filter that operates in a predictor-corrector manner. Combining bluetooth serial communication module gives the system a full mobility and makes the system affordable, lightweight, energy-efficient, and practical. Full 6DOF recovery and the full mobility of proposal system enable the user to interact with mobile device like PDA and provide the user with natural interface.

Keywords : Hybrid tracking system, Vision based tracking, Inertial tracking, Virtual Reality.

I. 서 론

가상 현실(Virtual Reality)은 컴퓨터가 만들어 놓은 환경 속에 3차원 공간을 사용자 마음대로 네비게이션(navigation)하여 새로운 경험을 갖게 하는 컴퓨터 그래픽 응용의 최첨단 기술을 말한다. 가상현실에서 가장 중요한 점은 사용자가 컴퓨터가 만들어 낸 가상세계와 3차원적인 상호 작용(3D interaction)을 할 수 있어야 한다는 것이다. 이러한 측면에서 보았을 때, 사용자에게 물리적 환경과 가상 환경 사이를 자연스럽게 이어주는 것은 매우 중요하다. 그래서 정확한 위치 추적(tracking)은 가상 환경을 더욱 유용하게 하는데 매우 필수적이다. 가상현실 어플리케이션을 위한 트래킹 시스템의 필수 조건들은 다음과 같다.

- 정밀한 3차원 위치와 자세 측정
- 떨림(jitter) 현상이 없거나 극히 작음
- 실시간 위치 추적
- 입력 센서로부터 시스템 출력까지의 지연현상이 없어야 함.
- 사용자 움직임에 제약이 없어야 함.

현재 가상 현실 어플리케이션을 위한 다양한 종류의 트래킹 시스템(자기장, 관성, 시각 기반(카메라), 초음파 등등)들이 있다. 하지만, 각각의 시스템은 하나 이상의 단점을 가지고 있으며, 결국은 시스템을 사용하는데 있어서 제약을 주게 된다.

그러면 두 가지 이상의 트래킹 시스템을 결합시킨 하이브리드 트래킹 시스템(hybrid tracking system)을 생각해 보자. 두 시스템을 결합시킴으로써 독립적으로 사용되었을 때보다 각 시스템이 갖고 있는 장점은 살리고 단점은 상호 보완할 수 있는 트래킹 시스템 구현이 가능해 질 것이다

본 논문에서는 시각 기반 트래킹 시스템에 관성 센서를 결합시킨 실용적인 하이브리드 트래킹 시스템을 제안하여 기존의 시각 기반 트래킹 시스템 구현에 있어서 나타나는 문제점을 해결하고 좀더 나아가 여러 가지 가상환경 어플리케이션으로의 응용을 다루고자 한다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 시각 기반 트래킹 시스템과 관성 센서를 이용한 트래킹 시스템(Inertial tracking system)의 제작 및 실험에 대해 소개하고 두 개의 트래킹 시스템을 결합시킨 관성과 시각 기반 하이브리드 트래킹에 대해 소개한다. III장에서는 제안한 시스템의 실험 및 가상 현실 어플리케이션의 몇 가지 예에 대해 설명하고, IV장에서 결론을 맺는다.

Ⅱ. 관성과 시각 기반 하이브리드 트래킹 시스템

1. 시스템 구성

제안하는 관성과 시각 기반 하이브리드 트래킹 시스템은 Pentium IV프로세서를 가진 일반 데스크 탑 PC 환경에서 구축되었고 4개의 고해상도(640X480) 영상을 실시간(30frame/sec)으로 처리하기 위해 2대의 컴퓨터를 이용하여 데이터를 분산처리 하였다. 적외선 방사체(IR emitter)와 적외선 카메라를 사용하여 가상 현실의 다양한 조명 환경을 극복하고 간단한 영상 처리만으로 특징점을 추출하여 연산량을 줄이고 실시간 트래킹을 보장하도록 하였다. 4개의 적외선 영상으로부터 특징점을 추출함으로써 1대 혹은 스테레오 카메라를 이용했을 때보다 정확성을 높이고 Occlusion현상에 의한 성능 저하를 최소화 하였다 .

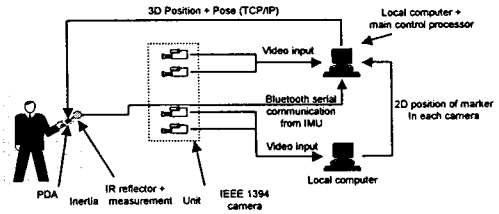


그림 1. 가상현실 어플리케이션을 위한 하이브리드 트래킹 시스템 구성.

2. 관성센서를 이용한 트래킹

관성 센서는 3개의 가속도계(accelerometer)와 3개의 자이로스코프(gyroscope)로 구성되어 있으며 센서로부터 나오는 신호를 A/D 변환(converting)을 하여 3축의 각속도 성분(roll, yaw, pitch)과 가속도 성분(a_x, a_y, a_z)을 얻는다. 관성 센서의 위치와 자세를 계산하기 위해 이산 칼만 필터(the Discrete Kalman filter)[1]를 이용하여 X, Y, Z축의 각속도와 가속도의 6개 신호의 정확한 값을 추정해보고, 그와 동시에 각 축에 대한 회전과 이동을 구하는 알고리즘을 설계한다. 이때 두 개의 칼만 필터를 설계하여 자세와 위치를 각각 구하여 연산의 효율을 높였다. 그림 2는 두 개의 칼만 필터를 이용한 트래킹 시스템의 모델링을 나타낸다.

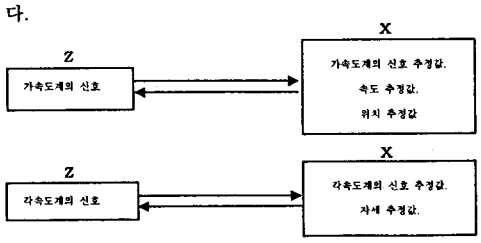


그림 2. 칼만 필터를 이용한 시스템 모델링.

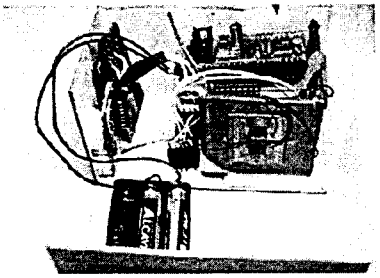


그림 3. 관성 센서를 이용한 트래킹 시스템의 구조 .

3. 시각 기반 트래킹

3.1 카메라 켈리브레이션(Camera calibration)

정확한 위치 트래킹을 위해서 카메라 켈리브레이션은 매우 중요하고 필수적인 작업이다. 4개의 입력 영상으로부터 오브젝트의 3차원 좌표를 재구성하기 위해서는 카메라의 내부 파라미터(intrinsic parameters)와 외부 파라미터(extrinsic parameters)를 구하여야 한다[2].

핀홀(pinhole) 카메라 모델에서는 3차원 공간상의 점들 $X_i = [X_i, Y_i, Z_i, 1]$와 2차원 영상에 대응되는 점들 $x_i = [u_i, v_i, 1]$는 다음과 같은 관계를 갖는다.

$$s\mathbf{x}_i = \mathbf{P}\mathbf{X}_i$$

본 논문에서는 위와 같이 영상과 3차원 공간과의 대응 관계를 이용하여 카메라 캘리브레이션 작업을 수행하였고 camera projective matrix P를 구하였다[3][4].

3.2 동기화(Synchronization)

4대의 카메라로부터 입력되는 고해상도 영상들을 실시간으로 처리하기 위해 카메라 입력 영상들을 2대의 컴퓨터에 나누어 분산처리 하였고 네트워크로 연결하여 처리된 데이터를 통신할 수 있도록 하였다. 정확한 트래킹을 위해서는 카메라들이 동시에 영상을 캡쳐(capture) 해야 하지만, 본 실험에서 사용된 저가형 카메라에서는 하드웨어적으로 동기화된 영상의 캡쳐가 불가능하다. 여기서는 이러한 문제를 해결하기 위해서 예측 동기화(predicted synchronization) 방법을 이용하였다. $m_1(n)$, $m_2(n)$, $m_3(n)$, $m_4(n)$를 각 카메라에서 n번째 입력 영상에서 특징점의 2차원 좌표라 하자

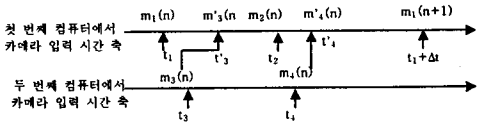


그림 4. 각 카메라 프레임 캡쳐 순서와 시간.

카메라 마다 프레임 캡쳐 시간이 그림 4에서와 같이

각각 다름을 알 수 있다. 네트워크를 통한 데이터 전송할 때도 어느 정도 시간이 소요되지만 특별한 네트워크 정체나 지연(delay)가 없고 데이터의 양이 크지 않으며 컴퓨터 간의 거리가 가까우면 데이터 전송 시간은 매우 짧아($\leq$1ms) $t_3 = t'_3$, $t_4 = t'_4$ 라 가정할 수 있다.

각각의 시간 t는 컴퓨터의 시스템 시간으로부터 구할 수 있으며 t'_4에서 i-번째(i = 1,2,3) 카메라의 특징점의 좌표 $m_i(n)|_{t=t'4}$ 를 예측하는데, 각 좌표의 4개의 이전 프레임에서의 좌표 값들을 가지고 Polyomial Extrapolation방법[5]을 이용한다.

동기를 맞추기 위해 제안한 방법이 완벽한 동기화를 이루지는 않지만 극히 짧은 시간(10ms 이내)에서 함수 $m_i(n)$은 선형성(linearity)을 만족하기 때문에 본 논문에서 제안하는 시스템 구현시 이용하였다.

3.3 3차원 좌표의 재구성(3D Reconstruction)

다수(적어도 2대 이상)의 카메라에서 추출한 특징점을 이용하여 3차원 좌표를 복원하는 과정은 다음과 같다. m(m$\geq$n)개의 카메라를 통해 대응되는 n개의 대응점이 다음과 주어졌다고 가정하자.

(u_1, v_1), (u_2, v_2), (u_3, v_3), (u_4, v_4), ... (u_n, v_n)

핀홀(Pinhall) 카메라 모델에 따르면

$$s_1[u_1, v_1, 1]^T = P_1[X, Y, Z, 1]^T \quad (1)$$

$$s_2[u_2, v_2, 1]^T = P_2[X, Y, Z, 1]^T \quad (2)$$

$$s_3[u_3, v_3, 1]^T = P_3[X, Y, Z, 1]^T \quad (3)$$

$$\vdots$$

$$s_n[u_n, v_n, 1]^T = P_n[X, Y, Z, 1]^T \quad (4)$$

p_i^j 를 3X4 행렬 P_i 의 j 번째 행이라고 하면 각각의 스칼라 항들은 다음과 같다.

$$s_j = \mathbf{p}_3^j \tilde{\mathbf{x}}$$

s_j 를 식 (1)에서 (4)까지 대입하여 정리하면 다음과 같은 행렬식을 구할 수 있다.

$$\mathbf{AX} = \begin{bmatrix} \mathbf{p}^1{}_1 \tilde{\mathbf{x}} - u_1 \mathbf{p}^1{}_3 \tilde{\mathbf{x}} \\ \mathbf{p}^1{}_2 \tilde{\mathbf{x}} - v_1 \mathbf{p}^1{}_3 \tilde{\mathbf{x}} \\ \mathbf{p}^2{}_1 \tilde{\mathbf{x}} - u_2 \mathbf{p}^2{}_3 \tilde{\mathbf{x}} \\ \mathbf{p}^2{}_1 \tilde{\mathbf{x}} - v_2 \mathbf{p}^2{}_3 \tilde{\mathbf{x}} \\ \mathbf{p}^3{}_1 \tilde{\mathbf{x}} - u_3 \mathbf{p}^3{}_3 \tilde{\mathbf{x}} \\ \vdots \\ \mathbf{p}^n{}_1 \tilde{\mathbf{x}} - u_n \mathbf{p}^n{}_3 \tilde{\mathbf{x}} \\ \mathbf{p}^n{}_1 \tilde{\mathbf{x}} - v_n \mathbf{p}^n{}_3 \tilde{\mathbf{x}} \end{bmatrix} = 0$$

위 방정식의 해는 A^TA의 최소 고유치(eigenvalue)에 해당하는 고유벡터(eigenvector)이고, 특이치 분해(Singular Value Decomposition)를 통해 얻어진다.

4. 하이브리드 트래킹

순수 시각 기반 트래킹 시스템은 떨림 현상이 적고 다른 시스템과 달리 위치 추적 값에 대한 오차가 누적되지 않지만 모든 상황에서 오브젝트의 자세를 재구성(reconstruction)하지는 못하며 컴퓨터의 연산 처리 능력과 카메라의 초당 읽어 들일 수 있는 프레임 수의 한계로 인해 시간 분해능(temporal resolution)이 높지 못하다.

이에 비해서 관성을 이용한 트래킹 시스템은 일단 오브젝트에 부착이 되고 위치와 자세의 초기치만 있으면 외부 입력이 필요 없는 그 자체만으로 완벽한 센서이다. 또한 제작된 시스템은 가볍고 전력 소모가 적으며, 크기가 작기 때문에 가상 현실 어플리케이션을 위한 트래킹 시스템으로 적합하다. 하지만 관성을 이용한 트래킹 시스템은 움직임이 느리거나 속도의 변화가 없을 경우 센서에서 나오는 신호 대 잡음비(Signal to Noise ratio)가 낮다. 이와 같은 현상은 자이로스코프보다는 가속도계에서 현저하게 일어나며 관성 센서가 갖고 있는 이러한 단점 때문에 시간이 흐를수록 위치 데이터의 오차가 증가하게 된다[6][7].

따라서 각각의 시스템이 갖고 있는 단점들을 극복하고 좀더 나은 위치 추적 시스템을 개발하기 위해 본 논문에서는 관성과 시각 기반 하이브리드 트래킹 시스템(Hybrid Inertial and Vision based tracking system)을 제안한다. 각각의 센서를 결합시킴으로써 제안한 하이브리드 트래킹 시스템은 강인(robustness)하고 시간 분해능이 높으며 정확할 뿐만 아니라 떨림 현상과 노이즈에 대한 영향을 최소화 하였다.

제안한 시스템의 트래킹은 칼만 필터를 통해 그림 5와 같이 예측 및 보정(prediction -correction) 방식으로 수행되어진다. 관성 센서의 데이터는 약 100Hz로 샘플링 되어진다.

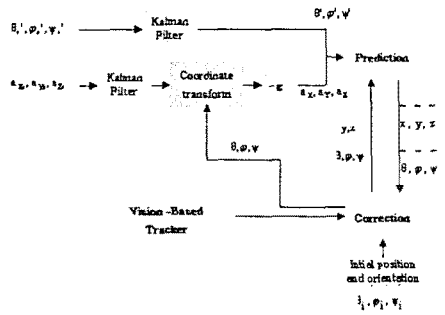


그림 5. 관성과 시각기반 하이브리드 트래킹 블록 다이어그램

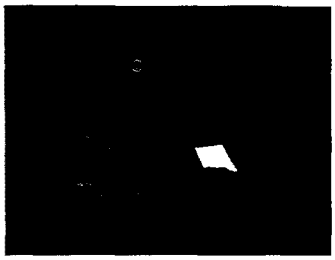
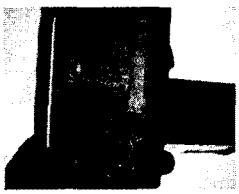


그림 6. Hybrid Tracker와 PDA를 이용한 Wand 어플리케이션

(a) 네비게이션 하는 사용자



(b) PDA를 통해 나오는 화면

(c) PDA를 통해 캡쳐된 영상

그림 7. Hybrid Tracker와 PDA를 이용한 가상 현실 네비게이션.

III. 실험 및 결과

본 논문에서는 제안한 하이브리드 트래킹 시스템을 이용하여 PDA를 이용한 가상 현실 어플리케이션들을 구현하였다. PDA에 제안한 하이브리드 트래킹 시스템을 부착시켜 사용자가 손에 쥔 PDA의 위치와 자세 정보를 추출하고 그것을 이용하여 사용자가 가상 현실과 인터렉션을 하는데 활용하였다. 시스템은 관성 센서로부터 입력되는 6개의 입력신호(3개의 가속도 성분 + 3개의 각속도 성분)로부터 PDA의 위치와 자세를 구한다. 또한 시스템은 블루투스 시리얼 통신을 이용하여 무선으로 트래킹을 수행하였으며, 4개의 적외선 입력 영상으로부터 구한 PDA의 위치 값을 가지고 관성 센서에서 구한 위치 값과 자세 값을 보정하여 시스템의 성능과 정확성을 높였다. 마지막으로 하이브리드 트래킹으로 얻은 3차원 위치와 자세 정보를 무선 랜(LAN)을 통해 PDA에 전달함으로써 PDA를 이용한 다양한 가상 현실 어플리케이션을 구현하였다. 그림 6는 가상 환경에서 PDA가 가리키는 방향으로 마치 우리가 지팡이(Wand)로 방향을 가리키듯이 가상의 포인팅(virtual pointing)을 하여 CAVE와 같은 몰입형 가상 공간에서 여러 가지 인터렉션을 할 수 있도록 구현되었다. 그림 7은 사용자가 PDA를 들고 움직이면서 PDA 화면을 통해 나오는 가상 현실을 직접 네비게이션하고 체험하도록 구현되었다. 하이브리드 트래킹 시스템으로부터 얻어지는 PDA의 위치와 자세에 따라 화면을 렌더링(Rendering)시킴으로써 사용자에게 사실감을 느낄 수 있도록 하였다.

IV. 결 론

본 논문에서는 다수의 카메라와 관성 센서를 이용하여 사용자나 오브젝트의 움직임을 추적할 수 있는 하이브리드 트래킹 시스템을 제안하였다. 관성 센서가 갖고 있는 완벽한 자세 복원 및 높은 시간 분해능(temporary resolution)과 외부 간섭에 영향을 받지 않는 강인함(robustness)은 빠른 움직임에서도 시스템의 움직임의 예측과 3차원 복원을 가능하게 하며, 시각 기반의 트래킹 시스템이 갖고 있는 정확성과 낮은 떨림 현상은 관성 센서가 갖고 있는 낮은 정확성을 보완하여 단순히 두개의 시스템을 결합시키는 것이 아니라, 서로의 장점을 살리고 각각의 시스템이 갖고 있는 단점을 보완할 수 있도록 하였다. Bluetooth 시리얼 통신을 이용하여 무선으로 트래킹 시스템으로부터 신호 입력을 가능케 함으로써 사용자가 가상 환경에서 전선(wire)에 의한 활동의 제약을 없게 하고 사용자로 하여금 자유로운 운동성을 보장하도록 하였다. 또한 크기와 무게를 줄여 사용자가 가상 현실에서 다양한 상호작용(interaction)을 할 수 있는 가상 현실 어플리케이션을 위한 실용적인 하이브리드 트래킹 시스템 개발이라는 목적에 부합하도록 하였다. .

참고문헌

[1] G. Welch and G. Bishop. " An Introduction to the Kalman Filter." *Technical Report, Department of Computer Science*, University of North Carolina, 2003.

[2] R. T. Tsai, "An Effective and Accurate Camera Calibration Technique for 3D Machine Vision," in *Proceedings CVPR'86*, pp.364-374, 1986.

[3] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press 2000.

[4] O. Faugeras, *Three-Dimensional Computer Vision : a Geometric Viewpoint*, MIT press, 1993.

[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipe in C," Cambridge University Press, 1988.

[6] Pinz, M. Brandner, H. Ganster, A. Kusej, P. Lang, and M. Ribo. "Hybrid tracking for augmented reality." *ÖGAI Journal*, 21(1):17-24, 2002.

[7] P. Lang, M. Ribo, and A. Pinz. " A new Combinational of Vision-Based and Inertial Tracking for Fully Mobile, Wearable, and Real-Time Operation." *26th Workshop of the AAPR/ÖGAM, OCG, Schriftenreihe, Band 160*, pp. 141-148, Graz, Austria, September 2002.