

도심환경에서 증강현실을 위한 카메라 기반 트래킹방법

*김 희 관, 이 명 호
(주)제니텀

A method of Single camera based tracking for mobile augmented reality system in an urban environment

*Albert Heekwan Kim, MyungHo Lee
ZENITUM

요약 : 증강현신 기술은 지리정보 가시화 영역에 높은 잠재력을 갖고 있다. 특히, 실제 환경에 가상의 지리정보를 증강시킴으로써 사용자가 직관적으로 정보를 습득, 사용할 수 있다는 점이 장점으로 작용할 것이다. 하지만 기존의 실외환경 증강현실 시스템들은 GPS를 사용하거나 RFID 또는 마커를 사용하여 사용자의 위치를 트래킹 하고자 하였다. 이러한 연구들은 도심환경과 같은 복잡한 공간에서 정확도가 떨어지게 되고, 많은 양의 센서를 설치해야 하기 때문에 실용성에 의문이 든다. 이에 본 논문은 GPS의 장점을 살리면서 별도의 기반시설을 필요로 하지 않는 새로운 트래킹 방법을 제안하고자 한다.

1. 서론

증강현실(Augmented reality, AR)기술은 가상현실(Virtual reality, VR)시스템에서 파생되어 나온 기술이다. 가상현실 시스템은 사용자에게 가상의 공간에 완전히 몰입할 수 있는 환경을 제공한다. 그 예로 사용자의 위치와 시점을 추적해 시점의 변화에 따른 3차원 영상을 제공해 주어 가상의 공간 안에 있는 느낌을 극대화 시킨다. 증강현실은 실제 환경을 기반으로 하는 가상현실 시스템으로 볼 수 있다. 증강현실 시스템은 카메라를 통해 들어오는 영상에 가상의 물체를 혼합해서 사용자에게 제공한다(Azuma, 1997). 사용자가 실제 공간 위에서 가상의 물체와 상호작용할 수 있기 때문에 증강현실 기술은 위치 기반 서비스(Location based service, LBS)에 적합하다. 증강현실이 지리정보시

스템(Geographic information system, GIS)에 적용된다면 사용자는 현재 바라보고 있는 건물의 배관파이프를 볼 수 있고, 마주하고 있는 산에 관한 정보를 실제 산과 함께 볼 수 있게 된다.

Mobilizy smartphone solutions의

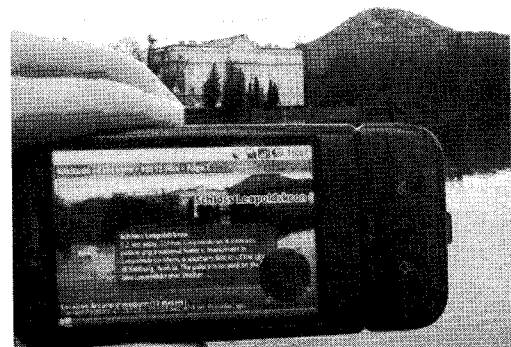


그림2. Wikitude. 카메라를 통해 보이는 건물에 설명을 증강하여 제공한다.



그림1. GIS에 사용된 증강현실

Wikitude는 GPS정보와 디지털 나침반을 사용하여 사용자의 위치를 기반으로 한 관광 정보를 제공한다. Wikitude는 GPS 데이터를 통해 사용자의 위치정보를 제공 받고, 나침반 정보를 이용해 사용자의 시점을 계산하여 카메라를 통해 들어오는 영상에 보이는 산이나 건물에 적절한 주석을 달아 준다(Mobilizy, 2008).

하지만 GPS정보를 이용하는 증강현실 시스템은 GPS 고유의 오차로 인해 위치정보의 정확도가 크게 떨어지게 된다. 도심과 같은 좁은 공간에 복잡한 정보를 갖는 환경에서 증강현실 시스템을 구현하려면 좀 더 정확한 위치 트래킹 방법이 필요하다. Föckler et al.(2005)는 모바일 폰을 이용한 박물관 안내 시스템 PhoneGuide를 개발하였다. PhoneGuide는 GPS정보와 같은 외부 장치의 지원 없이, 영상을 인식하고 그에 따른 적절한 정보를 사용자에게 제공한다. 하지만 인식하려는 영상정보를 미리 알고 있어야 하기 때문에 대량의 이미지가 제공되는 도심 환경에서 사용하기에는 어려움이 있다.

2. MoTRACK

MoTRACK은 카메라 기반 위치 추정 시스템과 센서 기반 위치 추정 시스템을 조합하여 도심과 같은 복잡한 환경에서도 매끄러운 증강현실을 제공할 수 있는 위치 추정 기법이다. MoTRACK은 멀티센서 트래커로서 GPS, 디지털 나침반센서, 가속도센서와 카메라를 사용한다. GPS 모듈을 통해 사용자의 Global position 정보를 얻고, 디지털 나침반 센서를 사용해 카메

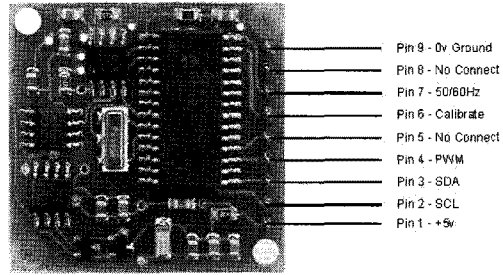


그림4. 디지털 나침반 모듈

라의 시점 정보를 얻는다. 또한 카메라에 정렬된 가속도센서를 통해 카메라의 자세를 추정하게 된다. 이렇게 센서를 통해 얻어진 1차 적인 정보를 이용해 대강의 위치와 자세를 추정하고, 카메라를 이용해 좀 더 세밀한 위치를 추정하게 된다.

3. 위치추정방법

3.1 GPS, 나침반, 가속도센서

GPS는 Global Positioning System으로 인공위성에서 보내오는 정보를 수신하여 위치를 측정할 수 있는 위치측정시스템이다. GPS 정밀도는 여러 오차 근원의 합에 의해 결정되며, 현재 널리 사용되는 GPS 오차는 약 20m정도 이다. MoTRACK은 GPS를 이용하여 어느 정도 오차를 갖는 사용자의 Global position을 정하게 된다. 이때 사용자는 GPS 모듈과 나침반센서, 가속도 센서 그리고 카메라를 포함하는 모바일 기기이다.

디지털 나침반 모듈 CMPS-03은 Philips KMZ51 지자기 센서로 구성되었다. 직각으로 배치된 두 개의 자기장 센서를 사용하여 지구 자기장의 수직 요소의 방향을 계산하여 출력값을 얻게 되며, I2C 인터페이스를 통해서 값을 읽는다. 디지털 나침반의 출력은 0~359.9의 값을 갖으며, 이는 북위를 기준으로 한 값이다. 사용자는 GPS값과 디지털 나침반 값을 사용하여 현재의 위치와 바라보고 있는 방향을 알 수 있게 된다.

가속도센서로는 LIS3LV02DQ를 사용하며 SPI 인터페이스를 통해서 값을 읽는다.

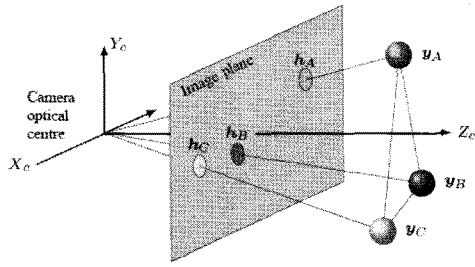


그림 5. P3P 알고리즘

정지상태의 물체에서 3축 가속도센서의 값은 중력가속도 값을 나타낸다. 이 값을 사용하여 카메라의 자세를 계산해 낼 수 있다.

카메라는 제한된 시야각을 갖고 있으므로, 앞서 계산된 위치, 방향, 자세를 이용해 카메라가 측정 가능한 관심 영역 (Region of interest)을 정의할 수 있게 된다. MoTRACK은 1차로 관성센서와 GPS 값을 통해 어느 정도 오차를 갖는 사용자의 위치를 추정한다고 하였다. 그 다음 단계는 사용자의 위치를 좀 더 세밀하고 추정하는 것이다. MoTRACK은 사용자의 세밀화 과정을 위해 영상 매칭을 사용한다. 이때 영상 매칭에 사용되는 특징점들은 데이터베이스에 미리 정의 되어 있다고 가정한다. 데이터베이스 내에 존재하는 모든 특징점들과 영상 매칭을 수행하는 것은 계산의 부하가 크기 때문에 앞서 설정한 관심 영역 내에 존재하는 특징점들과 영상 매칭을 수행하게 된다.

영상 매칭을 통해 위치를 계산하는 방법에는 호모그래피(homography)를 이용하는 방법과 P3P알고리즘(Perspective 3 point algorithm)이 있는데, MoTRACK은 P3P알고리즘을 이용한다.

3.2 Perspective 3 Point Algorithm

P3P알고리즘은 World coordinate 상에서 좌표를 알고 있는 3차원의 세 점들과 Image plane에 투사된 3점의 위치를 통해 카메라의 위치를 계산해 낸다.

그림에서 3개의 점 y_A, y_B, y_C 는 데이터베이스 내에 저장된 특징점이며 3차원 위

치를 알고 있는 점이다. Image plane에 h_A, h_B, h_C 는 현재 영상에서 NCC(Normalized cross correlation)을 통해 데이터베이스에 저장된 특징점과 매칭된 점들이다. Image plane에 투사된 점들을 통해 카메라 중심에서 각 점들로 뻗어가는 광선 사이의 각을 알 수 있고, 이 세 각 α, β, γ 와 각 점들 사이의 거리 a, b, c 를 이용하여 코사인 제 2 법칙을 적용하면 카메라와 각 점사이의 거리 s_1, s_2, s_3 에 대한 식을 3개 얻을 수 있다.

이 연립방정식의 해를 구하는 방법은 여러 가지가 있으나 MoTRACK에서는 Fischler 와 Bolles의 풀이 방법을 이용하였다(Fischler and Bolles, 1981).

3.3 MonoSLAM

매 영상 프레임마다 P3P알고리즘을 적용하는 것은 시스템에 부하를 주게 된다. MoTRACK은 실시간으로 동작하며 모바일 기기에 적합한 트래킹을 제공해야 하기 때문에 매 프레임마다 P3P알고리즘을 적용하는 대신 MonoSLAM알고리즘을 적용하여 사용자의 위치를 추적하였다. P3P알고리즘은 주기적으로 실행되어 MonoSLAM의 오차를 보정하도록 하였다(Williams et al., 2007).

MonoSLAM은 A. Davison에 의해 개발된 하나의 카메라만을 사용하여 카메라의 위치를 추적하고 주변의 지도를 생성하는 알고리즘이다. MoTRACK은 지도 생성부분을 제외한 트래킹 부분을 이용하였다(Davison et al., 2007).

4. 결론

본 논문에서는 멀티센서를 이용하여 카메라의 위치 추정하는 알고리즘을 제안하였다. 도심과 같은 빌딩이 많고, 트래킹 범위가 세밀해야 하는 공간에서는 기존의 GPS기반의 위치 추적을 통해 사용자에게 적절한 증강현실 서비스를 제공할 수 없다. 그렇기 때문에 본 논문은 카메라를 사

용하여 위치 추적의 정확도를 높일 수 있는 방법을 제안하였다.

전체 과정은 GPS와 관성센서를 사용하여 어느 정도 오차를 갖는 사용자의 위치를 추정하고, 추정된 위치를 통해 데이터베이스에서 영상 매칭에 필요한 특징점을 제공받아 3P3알고리즘으로 사용자의 정확한 위치를 추정하게 된다. 한번 위치를 추정한 뒤에는 한 동안은 MonoSLAM의 트래킹 기법을 이용하여 증강현실 서비스를 끊김 없이 제공한다.

이 논문에 사용된 영상 매칭 방법은 MonoSLAM에서 사용된 이미지 패치를 기반으로 구현되었다. 그렇기 때문에 향후 좀 더 위치추정의 안정성을 높이기 위해 SIFT(Scale Invariant Feature Transform) 또는 SURT(Speeded up Robust Feature)를 사용하는 방법을 모색해야 할 것이다.

이 논문에서 사용된 GPS, 디지털 나침반, 가속도센서 그리고 카메라는 최근 출시되는 모바일 장치에 기본적으로 장착되어 나오고 있기 때문에 본 알고리즘을 적용하기 위해 별도의 장치를 추가로 설치할 필요가 없어 유용할 것이라고 생각된다.

감사의글

본 연구는 건설교통부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과 제의 연구비지원(07 국토정보 C05)에 의해 수행되었습니다.

참고문헌

- Azuma, R.T., 1997. A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6, 4(August 1997): 355-385.
- Davison, A.J., I.D. Reid, N.D. Molton and O. Stasse, 2007. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 29(6): 1052-1067.
- Fischler, M.A. and R.C. Bolles, 1981. RANdom SAmple Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6): 381-395.
- Föckler, P., T. Zeidler, B. Brombach, E. Bruns, and O. Bimver, 2005. PhoneGuide: Museum guidance supported by on-device object recognition on mobile phones. Proc. International conference on mobile and ubiquitous computing: 3-10.
- Mobilizy smartphone solutions, 2008. <http://www.mobilizy.com>
- Williams, B., P. Smith and I.D. Reid, 2007, Proc. International Conference on Robotics and Automation.