

증강현실 기술 최근 동향 및 분석 - 3차원 객체 추적 기술을 중심으로 -

*박한훈 **박종일

*부경대학교, *한양대학교

hanhoon_park@pknu.ac.kr, jipark@hanyang.ac.kr

Recent Trends and Analysis on AR Technology

- Focused on 3D Object Tracking Methods -

*Hanhoon Park **Jong-Il Park

*Pukyong National University, **Hanyang University

요약

최근 증강현실(augmented reality) 기술이 큰 주목을 받고 있다. 스마트폰, HMD 등과 같은 모바일 기기의 성능 향상 및 위치 기반 서비스의 보편화로 인해 증강현실 기술의 유용성을 입증하는 다양한 응용 분야들이 소개되고 있다. 증강현실은 기본적으로 현실세계에 가상의 디지털 콘텐츠를 자연스럽게 병치하여 인간의 감각과 인식을 확장시키는 실감미디어 기술이다. 증강현실을 구현하기 위해서는 사용자나 현실세계의 상황을 파악하고 인지하기 위한 컴퓨터 비전 기술, 가상의 디지털 콘텐츠를 생성하고 렌더링하기 위한 컴퓨터 그래픽스 기술, 증강현실 콘텐츠와 상호작용하기 위한 상호작용 기술, 사용자 중심의 증강현실 프로그램 개발을 지원하기 위한 저작 기술 등 다양한 요소 기술들을 필요로 한다. 본 논문에서는 컴퓨터 비전 기술 중의 하나로 객체의 3차원 움직임(포즈 변화)을 추적하기 위한 방법들의 최신 동향을 간략하게 분석하고, 향후 발전 방향에 대해 전망해 본다.

1. 서론

포켓몬고(Pokemon Go)와 같은 게임의 등장과 함께 증강현실(augmented reality) 기술에 대한 대중들의 관심이 폭발적으로 증가했으며, 글로벌 기업들은 증강현실 관련 회사를 인수하거나 선도 기술 개발 및 확보를 위해 엄청난 규모의 투자를 진행하고 있다. 이미 증강현실 관련 상품 및 서비스들은 주변에서 쉽게 찾을 수 있으며 우리 생활을 전반적으로 변화시키고 있다.

증강현실은 기본적으로 현실세계에 가상의 디지털 콘텐츠를 자연스럽게 병치하여 인간의 감각과 인식을 확장시키는 실감미디어 기술이다. 증강현실을 구현하기 위해서는 사용자나 현실세계의 상황을 파악하고 인지하기 위한 컴퓨터 비전 기술, 가상의 디지털 콘텐츠를 생성하고 렌더링하기 위한 컴퓨터 그래픽스 기술, 증강현실 콘텐츠와 상호작용하기 위한 상호작용 기술, 사용자 중심의 증강현실 프로그램 개발을 지원하기 위한 저작 기술 등 다양한 요소 기술들을 필요로 한다. 증강현실의 가장 기본적이고 핵심적인 기술은 카메라 영상으로부터 객체(object)의 3차원 움직임(포즈 변화)을 추적하는 것이다. 본 논문에서는 컴퓨터 비전 기술 중의 하나로 객체의 3차원 움직임(포즈 변화)을 추적하기 위한 방법들의 최신 동향을 간략하게 분석하고, 향후 발전 방향에 대해 전망해 본다.

2. 객체 추적 기술의 개념 및 요소 기술

증강현실에서 객체 추적은 연속된 카메라 영상으로부터 영상을 촬영한 카메라나 관심 객체의 3차원 움직임, 또는 카메라와 관심 객체 사이의 상대적인 3차원 포즈 변환(회전 및 이동 행렬)을 추정하는 기술을 말한다. 이를 위해서는 객체의 3차원 좌표가 카메라 영상 좌표로 어떻게 투영되는지, 카메라 영상으로부터 객체의 3차원 좌표를 어떻게 계산해 낼 수 있는지 등을 알아야 한다. 즉, 카메라 투영(projection) 행렬 계산, 영상 특징(image feature) 검출(detection)/서술(description)/매칭(matching)/추적, 3차원 객체 모델링(modeling)과 같은 요소 기술들을 필요로 한다. 객체 추적과 관련된 연구는 주로 이러한 요소 기술들을 어떻게 제공하느냐에 따라 나누어질 수 있다.

3. 객체 추적 기술의 최근 동향

객체 추적 기술은 꾸준히 개발, 발전되어 왔으며, 높은 기술 성숙도로 인해 증강현실 응용 분야의 확대에 크게 기여하고 있다. 그러나, 여전히 해결하기 힘든 기술적 난제들이 있다. 대표적인 예로, 반짝이는 객체, 투명한 객체, 모양이 가변적인(deformable) 객체, 빠르게 움직이는 객체, 가려짐이 매우 심한 객체 등을 추적하는 것은 쉽지 않으며, 역광이나 야간 환경처럼 아주 밝거나 어두운 조명 환경에서 제약된 동적 범위(dynamic range)를 가진 카메라를 이용하여 객체를 추적하는 것은 매우 어려운 일이다. 또한, 최근 증강현실 구현을 위한 핵심 플랫폼인 HMD, 스마트폰, 태블릿과 같은 모바일 기기의 경우 컴퓨팅 리소

스가 부족하기 때문에 많은 계산량을 필요로 하는 정교한 객체 추적을 실시간으로 수행하는 것은 여전히 쉽지 않다. 결과적으로, 최근 발표되고 있는 객체 추적 기술의 대부분은 이러한 기술적 난제를 극복하는데 초점을 두고 있다(예, 그림 1 참조). 결국, 비전 기반 객체 추적 기술의 한계를 극복하기 위해 최근 스마트폰 등에 보편적으로 탑재되어 있는 GPS, IMU 등의 부가적인 센서를 함께 활용한 방법들이 다시 주목받고 있다.



그림 1. 미분 정보를 특정 벡터로 활용한 반짝이는 객체(사각 평면) 추적[1]. (좌) 밝기 정보 이용, (우) 가우시안 미분 정보 이용.

또한, 깊이 센서의 상용화 및 객체 추적에서의 깊이 정보의 유용성 검증, 딥러닝(deep learning) 기술의 발전 및 높은 잠재성으로 인해 최근 발표되고 있는 객체 추적 기술의 대부분은 깊이 센서와 딥러닝 기술 중의 하나 혹은 둘을 함께 활용하는 추세이다.

깊이 센서는 객체의 상태나 조명 환경에 의한 영향이 거의 없기 때문에, 객체의 가려짐, 열악한 조명 환경(낮은 조도, 극심한 조명 변화), 텍스처 정보 부족 등의 문제에 효과적으로 대응할 수 있다. 그러나, 기존의 ICP 기반의 방법은 local minima에 빠질 위험이 크고 계산 효율성도 떨어지기 때문에, 최근 연구들은 랜덤 포레스트나 CNN을 활용하여 보다 강건하고 효율적인 객체 추적이 가능한 방법들을 제안하고 있다[2, 3]. 특히, 템플릿 기반 추적 기술에서 서로 다른 포즈를 가지는 객체의 지역적인 특징 정보를 학습할 때, 깊이 정보를 함께 사용함으로써 가려짐에 효과적으로 대응할 수 있다는 것이 확인되고 있다[4].

딥러닝을 이용한 객체 추적은 광대한 양의 영상 데이터를 이용하여 객체 추적에 필요한 정보를 학습하기 때문에 기존의 객체 추적 기술이 직면해 왔던 다양한 문제들(복잡한 배경, 조명 변화, 모션 블러, 가려짐, 모양 변화 등)에 효과적으로 대처할 수 있다. 또한, 기존에는 영상으로부터 객체 추적에 적합한 특징 정보를 경험적, 수동적으로 결정했던 것과 달리, 딥러닝 기술을 활용하면 각 객체 추적에 적합한 최적의 전역적(global), 지역적(local) 특징 정보를 통계적, 자동적으로 학습할 수도 있다[4, 5].

4. 객체 추적 기술의 향후 전망

객체 추적 기술의 주요 활용 분야는 제조업, 산업 로봇, 에듀테인먼트 등인데, 대부분 텍스처가 적고 3차원 모델이 존재하는 객체를 대상으로 하기 때문에 객체의 메쉬(mesh), 템플릿(template), 깊이 정보를 함께 이용하는 것이 보편화되고 있으며, 앞으로 이러한 추세는 계속될 것으로 보인다.

객체 추적에 있어 깊이 정보의 유용성으로 인해 앞으로도 깊이 정보를 활용한 객체 추적 기술에 대한 연구는 큰 주류를 이룰 것으로 보인다. 다만, 실외나 모바일 환경에서 깊이 정보를 얻을 수 있는 방법에

대한 연구가 병행되어야 할 것이다. 깊이 카메라나 다시점 카메라를 이용하는 것은 실용성이 부족하기 때문에(특히 모바일 환경에서) 단일 시점 RGB 영상으로부터 깊이 정보를 얻을 수 있어야 하는데, CNN을 활용하여 깊이 정보를 획득하는 최근 연구[6]는 이를 위한 좋은 사례라 할 수 있다.

대부분의 객체 추적 기술은 시점에 따라 형태(깊이, 밝기/색상, 실루엣 등) 차이가 뚜렷한 작은 크기의 객체를 대상으로 한다. 이는 시점 간 객체 포즈의 모호성(pose ambiguity)을 줄이기 위한 것으로, 텍스처가 부족하고 모양이 단순하면서 큰 객체(전체를 한꺼번에 관찰할 수 없음)나 멀리 있는 객체의 경우 시점에 따라 형태의 차이가 작기 때문에 이러한 포즈의 모호성이 뚜렷해진다. 향후, 이러한 모호성을 가진 객체를 효과적으로 추적하기 위한 방법에 대한 연구가 필요할 것으로 판단된다.

PC 환경에서 검증된 객체 추적 기술들은 모바일 환경에서는 대부분 검증되지 않았으며 모바일 환경에서 구현할 경우 정확성, 강건성, 속도가 크게 떨어질 수 있다. 이는 컴퓨팅 리소스가 부족하기 때문에 PC 환경에 비해 간소화(simplification)되거나 근사화된 알고리즘이나 모델을 사용해서 구현하는 데서 비롯된다. 그러므로, 모바일 환경에서의 객체 추적 기술의 성능 개선은 앞으로도 꾸준히 진행되어야 할 중요한 연구 분야 중의 하나이다. 센서 융합은 모바일 환경에서 객체 추적 기술의 성능을 개선하기 위한 좋은 방안이 될 수 있다.

딥러닝 기술은 앞으로도 객체 추적을 위해 다양한 형태로 활용될 것이다. 다만, 딥러닝 기술을 적용하기 위해 필요한 광대한 학습 데이터를 보다 효과적으로 확보할 수 있는 방안이 필요하다. 한편, 학습 기반의 방법은 빠른 수렴을 통해 높은 강건성(robustness)을 가지지만 정확도가 높은 추적이 어려운 반면, 최적화 기반의 객체 추적 방법은 높은 정확도를 가지는 반면 local minima에 빠질 위험이 크기 때문에 강건성이 떨어진다[3]. 결과적으로 두 가지 방법론을 상호보완적으로 결합하는 것이 필요하다.

참고 문헌

- [1] A. Crivellaro and V. Lepetit, "Robust 3D Tracking with Descriptor Fields," Proc. of CVPR, 2014.
- [2] M. Garon and J.-F. Lalonde, "Deep 6-DOF Tracking," IEEE TVCG., vol. 23, no. 11, pp. 2410-2418, 2017.
- [3] D. J. Tan, N. Navab, and F. Tombari, "Looking Beyond the Simple Scenarios: Combining Learners and Optimizers in 3D Temporal Tracking," IEEE TVCG, vol. 23, no. 11, pp. 2399-2409, 2017.
- [4] W. Kehl, F. Milletari, F. Tombari, S. Ilic, and N. Navab, "Deep Learning of Local RGB-D Patches for 3D Object Detection and 6D Pose Estimation," Proc. of ECCV, vol. 3, pp. 205-220, 2016.
- [5] P. Wohlhart and V. Lepetit, "Learning Descriptors for Object Recognition and 3D Pose Estimation," Proc. of CVPR, 2015.
- [6] F. Liu, C. Shen, Guosheng Lin, and I. Reid, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields," IEEE TPAMI, vol. 38, no. 10, pp. 2024-2039, 2016.