동적 환경에 강인한 단안 카메라의 실시간 자세 추정 기법

박준형, 박인규
인하대학교
junhyeong2895@gmail.com pik@inha.ac.kr

Real-Time Monocular Camera Pose Estimation which is Robust to Dynamic Environment

Junhyeong Bak, In Kyu Park
Department of Electrical and Computer Engineering, Inha University

요약

증강현실이나 자율 주행, 드론 등의 기술에서 현재 위치와 시점을 파악하기 위해서는 실시간 카메라 자세 추정이 필요하다. 이를 위해 가장 일반적인 방식인 연속적인 단안 영상으로부터 카메라 자세를 추정하는 방식은 두 영상의 정적 객체 간에 견고한 특징점 매칭이 이루어져야한다. 하지만 일반적인 영상들은 다양한 이동 객체가 존재하는 동적 환경이므로 정적 객체만의 매칭을 보장하기 어렵다는 문제가 있다. 본 논문은 이 같은 동적 환경 문제를 해결하기 위해, 신경망 기반의 객체 분할 기법으로 영상 속 객체를 추출하고, 객체별 특징점 매칭 및 자세 추정 결과로 정적 객체를 특정해 매칭하는 방법을 제안한다. 또한, 제안하는 정적 객체 특정 방식에 적합한 신경망 기반 특징점 추출 방법을 사용하면 동적 환경에 보다 강인한 카메라 자세 추정이 가능함을 실험을 통해 확인한다.

1. 서론

실시간 카메라 자세 추정은 최근 각광받고 있는 증강현실이나 자율 주행, 드론 등의 기술에서 현재 위치와 시점을 파악하기 위해 필수적이다. 실시간 카메라 자세 추정 방식에는 센서를 통해 자세 정보를 바로 받아오거나[1], 단안 영상과 깊이 정보를 같이 사용하거나[2], 단안 영상만을 사용하는 방식[3] 등이 있다. 이 중 특수한 장치를 필요로 하지 않는 가장 일반적인 방식은 단안 영상만을 사용하는 방식이다. 연속적인 단안 영상으로부터 카메라 자세를 추정하기 위해서는 두 영상의 정적 객체 간에 견고한 특징점 매칭이 이루어져야 한다. 하지만 일반적인 영상들은 다양한 이동 객체가 존재하는 동적 환경이므로 정적 객체만의 매칭을 보장하기 어렵다는 문제가 있다.

본 논문은 이 같은 동적 환경 문제를 해결하기 위해, 신경망 기반객체 분할 기법으로 영상 속 객체를 추출하고, 객체별 특징점 매칭 및 자세 추정 결과로 정적 객체를 특정해 매칭하는 방법을 제안한다. 또한, 제안하는 정적 객체 특정 방식에 적합한 신경망 기반 특징점 추출 방법을 사용하면 동적 환경에 보다 강인한 카메라 자세 추정이 가능함을 궤도 분석 실험을 통해 확인한다.

2. 본론

연속되는 두 영상으로부터 카메라 자세를 추정하는 것은 두 영상의 정적 객체 간 특징점 매칭으로부터 5-points algorithm으로 essential matrix를 구하여 수행이 가능하다. 이 과정에서 동적 객체의 특징점과 같은 이상치가 매칭에 일부 포함되는 것은 RANSAC(random sample consensus)을 결합한 최소자승법을 사용해 해결이 가능하다. 하지만 영상 속에 다량의 이동객체가 큰 영역을 차지하는 경우는 RANSAC만으로는 해결이 불가능하다. 따라서 정적 객체만의 매칭을 보장하기 위해서는 매칭 수행 시 이동 객체에 해당되는 특징점을 사전에 완전히 제외할 수 있는 방식으로 구현이 필요하다.

이동 객체만을 특정해 특징점을 제거하기 위해서는 우선 객체 분할이 필요하다. 이를 위해 본 논문에서는 신경망 기반의 실시간 객체 분할 방법인 YOLACT[4]를 사용한다. YOLACT는 동일한 클래스의 객체들도 서로 다른 사물로 구분할 수 있는 방법이다. 이때, 객체 분할 결과가 실제 객체보다 작아 해당 하는 특징점을 포함하지 못하는 문제를 방지하기 위해 모폴로지 연산을 통한 영역 팽창 과정을 포함한다.

이와 같이 처리된 영상에서 객체로 분할되지 않은 배경영역은 온전히 정적인 객체로 간주되어 이동 객체 판정의 기준으로 사용된다. 분할된 객체들은 이동 객체를 특정하기 위해 개별적인 특징점 매칭 및 자세 추정을 수행한다. 이후 각 객체의 자세 추정 결과가 정적 객체에 대한 자세추정 결과와 비교 값 이상의 유사도를 가지면 정적 객체로 간주하며, 그 외의 객체는 전부 이동 객체로 간주하는 과정을 반복한다. 유사도는 L2 norm으로 계산된다. 정적 객체가 완전히 판정되면 모든 정적 객체의 특징점을 이용해 보다 정확한 최종적인 카메라 자세 추정 결과를 도출한다. 객체 분할과 이동 객체 특정 및 특징점 제외의 전 과정은 각

단계별로 처리된 영상과 함께 그림 1과 같이 나타내었다.

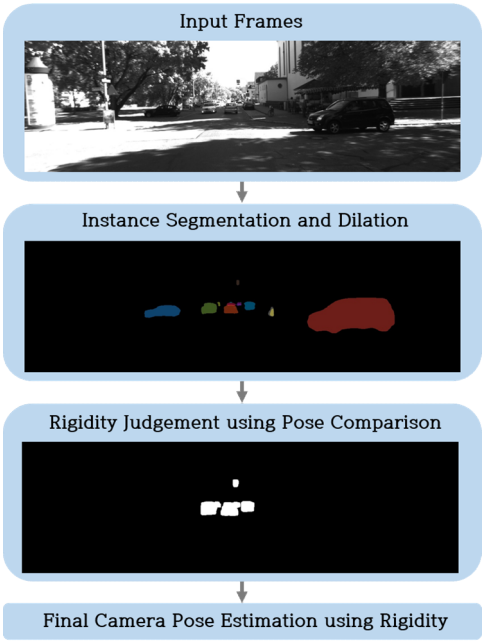

그림 1. 제안하는 정적 객체 특정 방법의 처리 과정

제안하는 방법은 고전적인 특징점 추출 및 매칭 기법을 사용해도 구현이 가능하다. 하지만 최근 연구에서 널리 사용되는 신경망 기반 특징점 추출 방법을 사용하면 보다 좋은 성능을 기대할 수 있다. 제안하는 방법은 이를 위해 Super point[5]를 사용한다. Super point는 256차원의 특징점을 구하기 때문에 보다 정밀한 매칭이 가능하다는 장점이 있다. 또한 영상 전체에 걸쳐 특징점이 분포하므로 이동 객체를 제외했을 때에도 여전히 많은 특징점이 남는다는 장점도 있다.

3. 실험 결과

제안하는 방법의 평가는 카메라 자세 추정 기반 궤도 추정 결과를 SIFT[6]기반의 고전적 기법과 정성적으로 비교해 수행했다. 평가에 사용한 데이터는 참값 궤도 정보를 제공하는 KITTI odometry dataset[7]을 사용했다. 비교 결과는 표 1과 같다. 신경망 기반 특징점 추출과 정적 객체 특정 방법을 같이 적용한 제안하는 방법은 고전적인 기법을 사용하는 방법보다 참 값 궤도와 유사함을 확인할 수 있다.

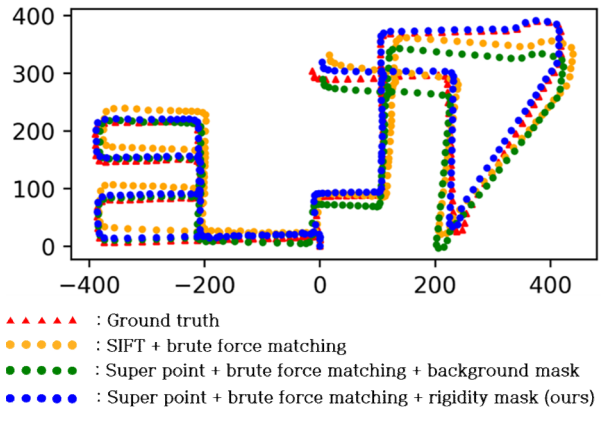

표 1. 카메라 자세 추정 기반 궤도 추정 결과의 정성적 비교

4. 결론

제안하는 방법은 가장 일반적인 환경이라고 할 수 있는 동적 환경에 강인한 단안 기반 카메라 자세 추정을 가능하게 했다는데 의의가 있다. 또한, 본 논문에서 사용된 신경망 기반 모델은 실시간 동작에 특화되어 있으면서도 고전적인 방법보다 우수한 카메라 자세 추정 결과를 보여주었다. 하지만 이동 객체의 크기가 너무 작거나 크고, 개수가 너무 많은 특수한 경우에는 좋은 성능을 발휘하기 어렵다는 한계도 있다. 이와 같은 제한 상황을 고려했을 때, 제안하는 방법은 주행 영상 같이 대체로 객체의 크기가 일정한 영상에서 실용적인 목적으로 충분히 사용 가능할 것으로 기대된다. 또한, 제안하는 방법을 발전시키면 모바일 기기에서도 동작 가능한 SLAM(Simultaneous Localization and Mapping)기반 AR이나, SfM(Structure from Motion)기반 3차원 복원 등에 적용이 가능할 것으로 보인다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00142, 스마트기기를 위한 온디바이스 지능형 정보처리 가속화 SW플랫폼 기술 개발)(2020-0-01389, 인공지능융합연구센터지원(인하대학교)).

참고문헌

[1] Z. Hu and K. Uchimura, ‘‘Fusion of vision, GPS and 3D gyro data in solving camera registration problem for direct visual navigation,’’ *International Journal of ITS*, vol. 4, no. 1, pp. 3-12, 2006.

[2] R. Mur-Artal and J. D. Tardos, “ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras,” *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255-1262, 2017.

[3] D. Burschka and E. Mair, “Direct pose estimation with a monocular camera,” In *Proceeding of International Workshop on Robot Vision*, 2008.

[4] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “YOLACT: Realtime instance segmentation,” In *Proceeding of IEEE International Conference of Computer Vision and Patter Recognition*, 2019.

[5] D. Detone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-Supervised Interest Point Detection and Description,” In *Proceeding of IEEE International Conference of Computer Vision and Patter Recognition workshop*, 2018.

[6] Y. Ke, R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” In *Proceeding of IEEE International Conference of Computer Vision and Patter Recognition*, 2004.

[7] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” In *Proceeding of IEEE International Conference of Computer Vision and Patter Recognition,* 2012.