

도심 환경의 다수 카메라에서 3차원 위치기반 물체 추적

3D Multi-camera Multi-object Tracking in Urban Spaces

최준혁¹·신현규¹·최해솔¹·이아현²·이강우²·최성록^{1†}

Jun Hyeok Choi¹, Hyun Gyi Hyeok Shin¹, Haesol Choi¹, Ahyun Lee², Kangwoo Lee²,
Sunglok Choi^{1†}

Abstract: This paper proposes multiple object tracking across multiple cameras. In contrast to previous researches on multiple object tracking for a single camera and re-identification, the proposed method investigates 3D visual geometry to derive 3D object location in the unit of meter, not pixel, with respect to the shared global coordinate. The 3D object location is a key clue to merge multiple object tracking generated from each camera. To fulfill this approach, we additionally developed multi-camera calibration, especially in the situation of little (or no) shared camera coverage. Moreover, we also develop 3D object localization on a general 3D plane. Finally, we demonstrated our 3D multi-camera multi-object tracking in the real urban space. We also provided quantitative evaluation of multi-camera calibration and 3D object localization

Keywords: Multiple Object Tracking, Multi-camera Multi-object Tracking, Camera Calibration

1. 서 론

도심 환경에는 많은 CCTV의 카메라가 설치되어 있고, 다수의 카메라 영상을 통해 사람과 물체, 그리고 환경의 변화를 관찰할 수 있다. 지능형영상감시(intelligent visual surveillance; IVS) 시스템을 비롯하여 이러한 다채널 대용량 카메라 영상에서 지능적인 방법으로 사람/사회에 필요한 유의미한 정보들을 도출하고자 하는 많은 의미있는 연구와 기술 개발이 있었다.

본 연구에서는 도심 환경에 설치된 다수의 CCTV 영상에서 사람과 차량과 같이 유의미한 물체들을 찾고 이들의 움직임을 추적한다. 특히 하나의 카메라에서 다수의 물체를 추적하는 DeepSORT[1]나 ByteTrack[2]과 같은 기존의 다개체추적(multiple object tracking; MOT) 기술을 확장하여 하나의 카메라가 아닌 다수의 카메라에서 다수의 물체를 추적하는 multi-camera multi-object tracking (MC-MOT) 문제를 풀고자

한다. 도심 공간의 다수의 카메라는 일반적으로 넓은 영역 커버하기 위해 공유되는 관찰 영역이 매우 제한적이거나 없는 특징이 있다. 이러한 특징 때문에 동일 물체의 여러 카메라에 중복으로 관찰되는 문제뿐만 아니라 사각지대에 따른 물체의 사라짐과 재출현 문제가 빈번하게 발생한다. 기존의 연구들[3]에서는 영상에서의 물체의 형상이나 특징에 의존한 동일 물체의 재검출(re-identification)을 통해 이를 해결하고자 하였다.

본 연구에서는 영상에서 물체의 위치에 대한 3차원적인 기하학적 해석을 통해 이러한 문제를 해결하고자 한다. 즉 픽셀 단위의 물체 위치를 파악하는 기존의 추적과 관련된 연구에서 벗어나 고정된 CCTV 카메라의 설치 위치와 방향을 고려하여 공간에서 하나의 전역좌표계를 기준으로 한 미터 단위의 3차원 위치를 도출한다. 이를 위해 카메라의 초점거리나 렌즈 왜곡계수, 그리고 설치 위치와 방향각에 대한 정보를 획득하는 1) 카메라 캘리브레이션 과정이 진행되어야 한다. 또 오르막이나 내리막, 다층구조와 같이 도심 공간에 존재하는 다수의 평면 위의 물체의 2) 3차원 위치를 도출하는 알고리즘이 필요하다. 마지막으로 3) 각 카메라에서 MOT을 통해 획득된 물체의 추적 결과를 전역좌표계를 기준으로 융합하는 기술이 개발되어야 한다. 본 논문에서는 위의 세 가지 연구 꼭지를 간략하게 소개하고, 실제 도심 환경에서의 다수의 CCTV 영상에서의 실험 결과를 통해 제안하는 기술을 정량적으로 분석·고찰한다.

※ This work was supported by an Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [22ZR1210, Development of Technology to reproduce and analyze the real-world cities into digitalized intelligent urban spaces].

1. Computer Science and Engineering Department, Seoul National University of Science and Technology (SEOULTECH), Seoul, Korea

2. Intelligent Robotics Research Division, ETRI, Daejeon, Korea

† Corresponding author (sunglok@seoultech.ac.kr)



[Fig. 1] Multi-camera multi-object tracking and its visualization on the satellite image

2. 다수 카메라의 3차원 물체 추적 및 실험

2.1 도로마커를 이용한 카메라 캘리브레이션 및 실험

도시 환경에서의 다수의 CCTV 카메라는 공통으로 관찰되는 영역이 매우 적기 때문에 일반적인 멀티-카메라 캘리브레이션 기술을 적용하기 힘들다.

본 연구에서는 일반적으로 도로에 존재하는 도로표식(road marker)과 위성영상(satellite image)을 이용하는 멀티-카메라 캘리브레이션을 사용하였다. 이를 통해 카메라의 초점거리나 렌즈 왜곡계수와 같은 내부 파라미터와 카메라의 위치와 방향 각과 같은 외부 파라미터를 모두 얻을 수 있었다. 위성영상은 CCTV 영상과 공유되는 영역을 갖고 있고, 해당 도로표식의 XY 위치 정보를 제공한다.

4개의 영상과 44개의 도로표식을 이용한 실험에서 초점거리 f 와 왜곡계수 k_1 만 고려한 카메라 모델이 교차점중의 재투영 에러가 15.14 픽셀로 도출되어 가장 적합한 모델로 선택하였다. 보다 복잡한 카메라 모델은 재투영 에러가 각각 21.73, 18.41, 24.04 픽셀이 도출되었다.

2.2 3차원 평면 위의 물체의 3차원 위치 도출 및 실험

3차원 공간에서의 카메라의 투영 관계를 이용하면 카메라에서 주어진 한 점이 특정 3차원 평면 위에 있을 때의 3차원 위치를 계산할 수 있다. 이는 광선(ray), 즉 3차원 벡터와 3차원 평면의 교점을 찾는 문제이다. 한 점은 카메라좌표계에서 표현되고, 카메라의 R과 t를 이용하여 해당 점의 광원의 시작점과 벡터는 각각 $-R^T$ 와 $R^T \hat{x}$ 로 표현된다. (\hat{x} 는 위치를 계산하고자 주어진 영상 위의 점에 왜곡을 보정하고 정규화된 이미지 평면에 표현한 점이다.)

앞서 실험에 사용된 4개 영상과 44개의 도로표식에서 각 도로표식의 재투영 에러는 평균 1.24 미터였다. 카메라와의 거리 20 미터 이내에서는 평균 0.54 미터의 에러가 있었고, 20 미터 이후에는 1.80 미터의 에러가 있었다.

2.3 다수의 카메라에서 MOT 결과 융합 및 실험

다수의 카메라의 MOT 결과는 위의 2.2장에서 도출한 물체의 3차원 위치를 이용하였다. 즉 사전에 등록된 물체와 새로 관찰된 물체의 종류(class)가 동일하고, 두 물체 사이의 직선거리가 정해진 임계값 이내일 때 같은 물체로 판단하였다. 또한 어떠한 사전에 등록된 물체와도 가깝지 않다면 새로운 물체로 등록하였다. 또 잠시 사라지대를 지나 관찰되지 않는 물체를 고려하여 등록된 물체에 나이(age) 개념을 도입하여, 관찰되지 않는 물체는 나이가 1씩 증가하도록 하였다.

약 200개의 물체가 등장하는 20분 분량의 4개의 영상에서 제안하는 MC-MOT를 적용하여 정성적으로 평가할 수 있었다. 각 영상에 YOLO v5를 적용해 물체를 검출하고 DeepSORT를 통해 MOT 결과를 얻었다. 기본적으로 잘 동작하였지만, 먼 거리의 물체의 3차원 위치의 부정확성(uncertainty)에 의해 고정된 임계값을 사용하는 것에 대한 문제 상황을 확인할 수 있었다.

3. 결론

본 논문에서는 다수의 카메라에서 다수 물체의 3차원 위치를 이용한 추적 기술을 소개하였다. 본 기술을 위해 도로표식을 이용한 멀티-카메라 캘리브레이션을 수행하였고, 광원의 벡터와 평면의 교점을 찾는 방법을 통해 물체의 3차원 위치를 추정하였다. 또 앞서 도출한 3차원 위치를 이용한 다수의 카메라에서 MOT 결과를 융합하였다.

References

- [1] Wojke et al., "Simple Online and Realtime Tracking with a Deep Association Metric", ICIP 2017
- [2] Zhang et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box", ECCV, 2022
- [3] He et al., "TransReID: Transformer-Based Object Re-Identification", ICCV, 2021