고속 비디오 안정화 알고리즘 기반 객체 탐지 성능 개선

최은우, 민병석

세종대학교

eunwooo0717@sju.ac.kr, bmin@sejong.ac.kr

Improvement of Object Detection with Video Stabilization

Eunwoo Choi, Byungseok Min*

Sejong University.

요 약

본 논문은 흔들림이 심한 비디오 환경에서 객체 탐지 성능을 개선하기 위한 고속 비디오 안정화 파이프라인을 제안한다. 기존 비디오 안정화 방식은 고해상도 연산으로 인한 병목과 CPU↔GPU 메모리 복사(Host↔Device transfer)로 인해 실시간 처리에 한계가 있었다. 본 연구에서는 저해상도 모션 추정, 고해상도 좌표 스케일링, GPU 기반 보정, 그리고 탐지기(YOLOv8m) 통합으로 구성된 최적화 파이프라인을 설계하였다. 실험 결과, 640×480 해상도에서 10fps 입력 스트림을 실시간 처리(프레임 예산 100ms)하면서 평균 약 30 ms/frame의 처리 시간을 기록하였고, 흔들림 환경에서 탐지 일관성이 향상됨을 확인하였다.

Ⅰ. 서론

실시간 객체 탐지 시스템은 교통·보안·스마트시티 등 다양한 응용에서 핵심 요소이며, 특히 CCTV 기반 분석에서는 장시간 스트리밍 환경에서의 안정적인 성능 유지가 중요하다. 그러나 실제 설치 환경에서는 바람, 진동, 지지 구조물의 탄성 등에 의해 프레임 간 전역 카메라 움직임이 지속적으로 발생한다. 이러한 전역 흔들림은 탐지 입력의 시각적 일관성을 저하시켜 탐지 누락 및 오검출을 유발할 뿐 아니라, 후단 다중 객체 추적기의 연관 성능을 저하시키는 주요 요인으로 작용한다. 비디오 안정화 알고리즘은 이를 개선하기 위한 전처리 방법으로 사용되는데, 일반적으로 광류(optical flow) 기반 변환을 추정한 뒤 프레임을 보정하는 방식으로 수행된다. 하지만 이를 객체 탐지기와 연계하여 사용하기 위해서는, 객체 탐지기가 GPU에서 동작하는 환경에서는 CPU 기반 워핑과 Host↔Device 메모리 전송이 반복되며 지연이 크게 증가할 수 있다. 특히 "CPU에서 보정 후 GPU로 재전송하여 탐지"하는 구조는 실시간 파이프라인의 병목을 유발하므로, 실제 다채널 실시간 시스템에서는 적용에 한계가 있어, 프레임 안정화 기술과 이를 위한 고속 처리 시스템 기술이 요구된다.

Ⅱ. 본론

카메라가 흔들리는 상황에서 객체 탐지 성능을 개선하기 위한 제안 방법은 그림 1과 같이 비디오 안정화 - 객체 탐지(및 추적) 통합 파이프라인이 되도록 구성하고, 입력 스트리밍 영상에 대해 저해상도 기반 모션 추정과 고해상도 보정을 결합하고, 보정과 탐지 과정을 GPU 상에서 연속적으로 수행되도록 한다.

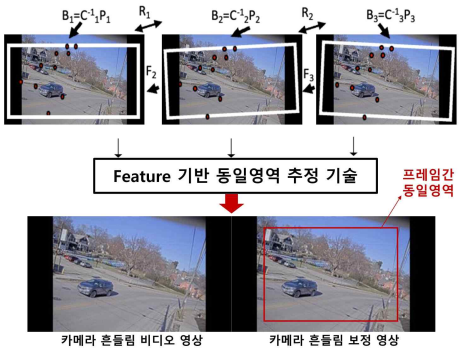

그림 1 특징점 기반 동일영역 추정 예시

구체적으로 입력 영상은 디코딩 이후 모션 추정을 위해 저해상도로 다운샘플링되며, PWC-Net[1] 또는 GlobalFlowNet(GFN)[2]을 이용하여 프레임 간 움직임을 추정한다. 추정 결과는 전역 Affine 변환으로 근사하여 요약되며, 누적 카메라 경로는 짧은 시간 윈도우에서 평활화되어 고주파 흔들림 성분이 제거된다. 이후 저해상도에서 얻은 변환은 고해상도 좌표계에 맞도록 스케일 보정되어 보정 변환으로 사용된다. 보정(워핑)은 GPU에서 수행되며, 안정화된 프레임은 YOLOv8m[3] 또는 RT-DETR[4]에 직접 입력된다. 프레임별 객체 탐지 성능을 개선하기 위하여, ByteTrack[5] 또는 BoT-SORT[6]와 같은 객체 추적 알고리즘을 결합하여 통합 객체 탐지 모듈을 구성한다.

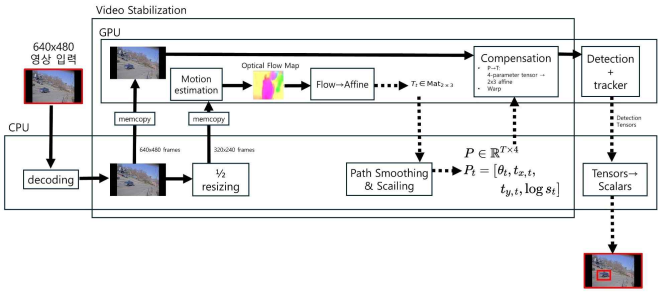

그림 2 제안 파이프라인 개요

제안 시스템은 입력 영상 프레임 및 비디오 안정화를 위한 모든 영상 프레임을 비디오 메모리에 유지함으로써 CPU 기반 워핑과 Host↔Device 메모리 전송 오버헤드를 최소화한다. 또한 저해상도 추정을 통해 추정 비용을 절감하고, GPU 배치 워핑으로 커널 호출 및 런타임 오버헤드를 감소시켜 실시간 처리에 적합한 구조가 되도록 전체 시스템 파이프라인을 구성한다.

전역 변환 추정 및 안정화

프레임 t에서 t+1로의 전역 변환을 A_t로 정의하면, 시점 t까지의 누적 카메라 경로는 다음과 같이 표현된다.

$$C_t = \prod_{(i=0)}^{(t-1)} A_i = A_{(t-1)} A_{(t-2)} \cdots A_0$$

비디오 안정화의 목적은 원래의 카메라 경로 C_t를 시간적으로 더 부드러운 경로 $\widetilde{C}_t$로 대체하는 것으로, 각 프레임에 적용되는 보정 변환은 다음과 같이 정의된다.

$$W_t = \widetilde{C}_t C_t^{-1}$$

본 논문에서는 C_t를 평행이동 성분 p_t과 회전 성분 θ_t 로 분해한 뒤, 각각에 대해 반경 r의 짧은 시간 윈도우에서 이동 평균 필터를 적용하여 부드러운 경로 $\widetilde{C}_t$를 생성한다.

$$\tilde{p}_t = \frac{1}{(2r+1)} \sum_{k=-r}^{r} p_{(t+k)} \qquad \tilde{\theta}_t = \frac{1}{2r+1} \sum_{k=-r}^{r} \theta_{(t+k)}$$

이와 같은 단순한 평활화 기반 접근법은 계산 복잡도가 낮아 실시간 스트리밍 환경에서의 비디오 안정화에 특히 적합하다.

전역 모션 추정의 연산량을 줄이기 위해 변환 추정은 저해상도에서 수행한다. 다만 저해상도에서 추정된 변환을 고해상도 프레임에 적용하기 위해서는 해상도 차이에 따른 좌표계 변환이 필요하다. 고해상도와 저해상도 간 스케일 비를 (S_x, S_y)라 할 때, 저해상도 좌표계에서 추정된 변환을 고해상도 좌표계로 변환하여 평행이동 성분이 픽셀 단위에서 일관되게 반영되도록 보정한다. 이러한 스케일 보정은 추정부 해상도를 낮추면서도 보정부 해상도에서의 보정 결과를 유지하기 위한 핵심 절차이다.

CPU 기반 Affine 보정블럭은 프레임 단위 반복 호출과 메모리 이동으로 인해 지연이 누적될 수 있다. 이에 본 연구에서는 GPU에서 affine grid 생성과 grid_sample을 이용해 워핑을 수행하고, 다수 프레임을 배치 단위로 처리하여 커널 호출 오버헤드를 감소시킨다. 또한 안정화 결과를 즉시 탐지기에 입력함으로써 안정화 - 탐지 - 추적 구간에서 Host↔Device 전송을 최소화한다. 이를 통해 워핑 단계의 지연을 완화하고, 전체 파이프라인의 병목을 변환 추정 및 탐지 단계로 정리하여 효율적인 성능 최적화를 가능하게 한다.

실험 및 결과

그림 3은 변위량(low/mid/high) 조건에서 PWC 및 GFN과 추정부 해상도(640×480, 320×240, 160×120)에 따른 안정화 성능을 비교한 결과를 제시한다. 여기서 추정 시간은 모션 추정에 소요된 평균 처리 시간이며, 보정 시간은 640×480 기준 워핑 적용 시간을 의미한다. 또한 SSIM을 통해 안정화 결과의 화질 및 프레임 간 구조적 유사도를 평가하였다. 본 실험은 해상도 축소에 따른 추정 시간 절감 효과와 보정 비용, 그리고 이에 따른 SSIM 변화를 함께 분석하는 데 목적이 있다. 실험은 640×480, 10 fps 입력에서 100프레임을 대상으로 GPU(A5000) 환경에서 수행하였다.

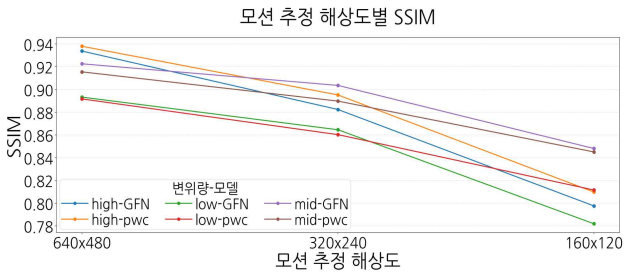

그림 3 변위량과 추정부 해상도에따른 SSIM 비교

그림 4는 입력 영상 유형과 탐지기 - 추적기 조합에 따른 객체 탐지 성능을 비교한 결과를 제시한다. 입력 영상은 원본, 인위적 흔들림 영상, 그리고 서로 다른 해상도에서 추정된 아핀 변환을 기반으로 한 안정화 결과(640 from 640/320/160 affine)로 구성된다. "640 from 320(160) affine"은 각각 320×240(160×120)에서 변환을 추정한 뒤 640×480 좌표계로 스케일 보정 하여 적용한 결과를 의미한다

.

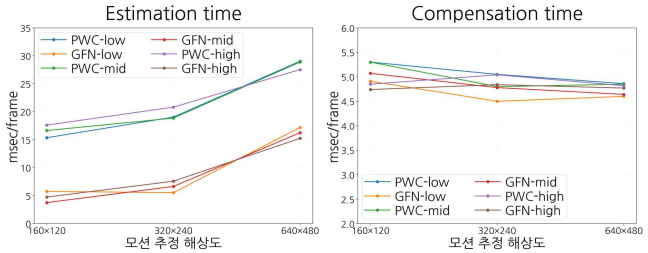

그림 4 모델·변위량별 추정부 해상도에 따른 시간 비교

탐지기는 YOLOv8m[3]과 RT-DETR[4]을 사용하였고, 추적기는 미적용(None), ByteTrack[5], BoT-SORT[6]를 고려하였다. 평가지표는 mAP@0.50과 처리 시간(ms/frame)이며, 안정화 적용 및 추적기 결합에 따른 정확도 - 속도 변화를 정량적으로 비교한다.

그림 5는 이러한 설정에서 측정된 mAP@0.50 결과를 시각적으로 요약하여, 흔들림 영상 대비 안정화 적용 시 탐지 성능이 전반적으로 개선됨을 보여준다. 특히 저해상도 추정 기반 안정화(640 from 320/160 affine)가 계산량을 줄이면서도 mAP을 유지 또는 개선하는지를 중점적으로 분석한다.

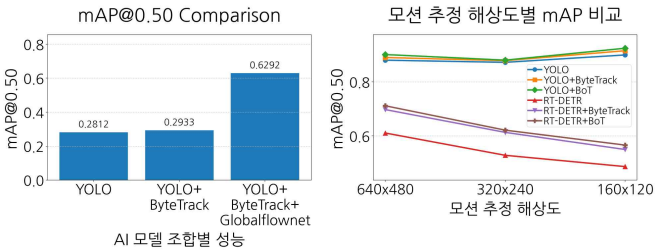

그림 5 비디오 안정화 적용에 따른 객체 탐지 성능 비교

Ⅲ. 결론

본 논문은 실시간 객체 탐지 시스템에서 비디오 안정화가 갖는 병목(특히 CPU 워프 및 Host↔Device 메모리 왕복)을 해결하기 위해, 저해상도 추정 - 고해상도 보정 스케일링 - GPU 배치 워프 - YOLO 탐지 통합 구조를 제안하였다. 표 1의 시간 분해 결과는 제안 방식이 워프 병목을 줄여 약 30 ms/frame 수준의 근접 실시간성을 달성함을 보여준다. 또한 표 2의 탐지 성능 비교를 통해 안정화 적용이 탐지 성능 개선에 기여할 수 있음을 정량적으로 분석하였다. 향후 연구로는 GPU 상에서의 스무딩/누적 경로 계산 완전 이식, 안정화 강도 적응, 그리고 탐지 - 추적 종단 지표(IDF1, MOTA 등) 기반의 최적화가 남아 있다.

본 논문은 실시간 객체 탐지 파이프라인에서 CPU 기반 워핑 및 Host↔Device 메모리 전송으로 인해 발생하는 지연을 완화하기 위해, 저해상도 전역 변환 추정 - 고해상도 스케일 보정 - GPU 배치 워핑 - 탐지(및 추적) 통합 구조를 제안하였다. 실험 결과, 제안 기법은 640×480, 10 fps 입력에서 평균 약 30 ms/frame 내외의 처리 시간을 달성하여 실시간 처리 가능성을 확인하였다. 또한 흔들림 영상에 대해 안정화 적용 시 탐지 성능(mAP)과 결과 일관성이 개선됨을 정량적으로 확인하였고, 악천후 환경과 같이 카메라가 흔들리는 상황에서 객체 탐지 성능을 향상하는 부분에 활용 기대된다.

ACKNOWLEDGMENT

이 논문은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 첨단분야 혁신융합대학사업의 연구결과입니다.

참 고 문 헌

[1] Sun, D. et al., "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume," CVPR, 2018.

[2] Wu, et al., "GlobalFlowNet: Learning Global Camera Motions for Unsupervised Video Stabilization," CVPR, 2022.

[3] Ultralytics, "YOLOv8: Next-Generation YOLO Object Detection," Tech Report/GitHub, 2023.

[4] Liu, et al., "RT-DETR: Real-Time End-to-End Object Detection with Transformers," ICCV, 2023.

[5] Zhang, et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," ECCV, 2022.

[6] Aharon, et al., "BoT-SORT: Robust Associations for Multi-Pedestrian Tracking," CVPRW, 2022.