

# 자율 주행을 위한 이벤트 데이터와 RGB 프레임의 특징점 기반 정합 프레임워크

주대연 / 광운대학교 영상처리시스템연구실(IPSL)

최근 자율 주행 환경에서 이벤트 카메라와 RGB 카메라를 융합(fusion)한 객체 탐지 기법에 대한 연구가 활발히 이루어지고 있다. 두 카메라의 융합을 통해 고속 객체 탐지 및 추적 성능을 향상시킬 수 있으나, 두 카메라로부터 획득된 데이터는 시간적·공간적으로 정합되어 있어야 한다. 현재까지의 대부분의 연구는 두 카메라 간의 정합이 이미 이루어진 상태를 가정하고 진행되었으며, 정합 문제 자체에 대한 연구는 부족한 상황이다. 본 논문은 자율 주행 상황에서 이벤트-RGB 융합 객체 탐지를 수행하기 위한 이벤트 카메라와 RGB 카메라 간 특징점 기반 정합 프레임워크를 제안한다. 제안하는 프레임워크는 이벤트 데이터를 누적하여 이벤트 프레임으로 변환하고, 이벤트 카메라의 특성을 모델링하여 RGB 영상으로부터 RGB 이벤트 프레임을 생성한다. 이후, 두 이벤트 프레임 간 특징점을 기반으로 호모그래피(homography) 변환 행렬을 유도한 뒤, 해당 행렬을 활용해 RGB 영상과 이벤트 프레임 간 정합을 수행한다. 제안하는 프레임워크는 자율 주행 환경에서 이벤트 카메라와 RGB 카메라로 구성된 스테레오 데이터 셋인 DSEC Dataset을 이용해 평가하였으며, 다중 객체가 서로 다른 깊이에 존재하는 복잡한 장면에서 정합

이 성공적으로 수행되었음을 주관적으로 확인하였고, 정합 전 평균 RMSE(Root Mean Square Error) 57.68에서 정합 후 평균 RMSE 24.46으로 객관적으로 우수한 성능이 입증되었다.

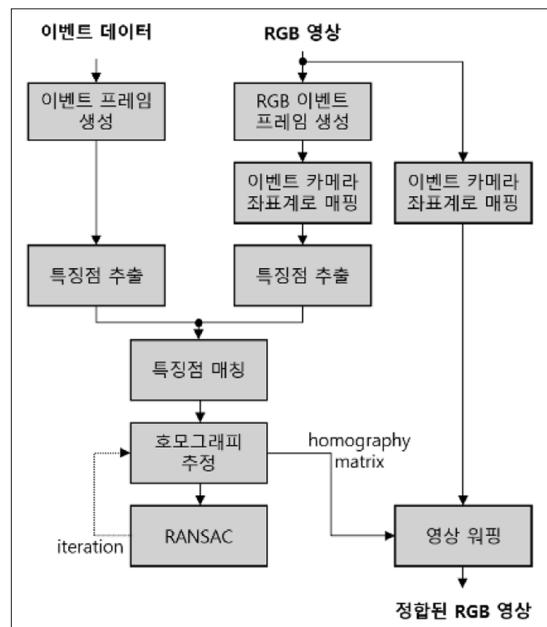
이벤트 카메라는 고정된 프레임율로 동기적으로 영상을 획득하는 기존 프레임 기반의 카메라와 달리, 각 픽셀이 독립적으로 작동하며 빛의 변화가 감지될 때마다 비동기적으로 이벤트를 출력하는 카메라이다. 이벤트 카메라는 기존 프레임 기반의 카메라 대비 높은 시간 해상도(high temporal resolution), 높은 동적 범위(high dynamic range)를 가지며, 저조도 환경 및 급격한 조도 변화에 강인하지만 밝기 변화가 감지된 픽셀에서만 데이터를 출력하기 때문에 출력 데이터가 희소(sparse)하고 색상과 같은 세부적인 시각 정보를 제공하지 못하는 한계가 존재한다. 반면, RGB 카메라는 밀집(dense) 데이터와 색상 정보를 제공하여 객체의 세부적인 구조와 특징을 명확하게 포착할 수 있지만 시간 해상도가 제한적이고 극단적인 조명 변화에 취약하다는 한계가 존재한다. 따라서 두 카메라의 융합을 통해 이벤트 카메라의 높은 시간 해상도와 동적 범위, RGB 카메라의 풍부한 색상 정보와

세부 구조 포착 능력을 결합하고 자율 주행 시스템의 안정성과 효율성을 효과적으로 향상시킬 수 있다. 그러나, 두 카메라를 효과적으로 융합하여 활용하기 위해서 이벤트 데이터와 RGB 데이터 간의 좌표계를 일치시키는 정합이 필수적이다.

정합은 서로 다른 위치와 방향 또는 센서로 촬영된 두 개 이상의 영상을 일관된 좌표계로 표현하는 과정이다. 의료 영상, 자율 주행 등에서 활용되며, 정확한 정합이 후속 처리의 성능을 좌우하기 때문에 중요한 역할을 한다. 전통적인 정합 알고리즘은 서로 다른 위치에서 취득된 두 개 이상의 RGB 영상에 대해 수행되었으며, 일반적으로 특징 점 추출, 특징 점 매칭, 변환 행렬 유도, 영상 변환의 네 가지 단계로 이루어진다. 특징점 추출 단계에서 영상 내 고유한 점, 패턴, 구조적 요소를 찾고 이를 표현하는 특징 기술자(descriptor) 벡터를 생성한다. 특징점 매칭 단계에서 두 개 이상의 영상에서 검출된 특징점과 해당 기술자를 비교하여 서로 대응되는 특징점들을 찾는다. 변환 행렬 유도 단계에서는 매칭된 대응점들을 이용해 영상 간의 기하학적 관계를 설명하는 변환 행렬을 계산한다. 마지막으로, 영상 변환 단계에서는 유도된 변환 행렬을 기반으로 영상의 좌표계를 변환하여 두 영상 간의 정합을 수행한다. 최근에는 다중 센서 정합에 관한 연구가 이뤄지고 있으며, 다중 센서 정합은 RGB 카메라, 적외선 카메라, LiDAR(Light Detection and Ranging), 이벤트 카메라 등 서로 다른 센서를 통해 획득된 데이터를 융합하여 일관된 좌표계로 표현하는 기술이다. 다중 센서 정합은 RGB 영상 정합과 유사하게 특징점 추출, 특징점 매칭, 변환 행렬 유도, 영상 변환 네 단계로 이루어진다. 하지만 서로 다른 센서로부터 취득된 데이터는 동일한 장면을 촬영 하더라도 정보를 표현하는 방식이 다르고 초점 거리(focal length), 픽셀의 실제 물리적 크기 등이 다르기 때문에 정확한 특징점 추출과 매칭이 어렵다. 특히 이벤트 카메라는 비동기적으로 빛의 변화를 감지하여 희소한 이벤트 데이터

터를 출력하는 반면 RGB 카메라는 고정 프레임율로 빛의 세기와 색상 정보를 밀집된 픽셀 단위로 표현하기 때문에 두 데이터 간 정합은 상당히 어려운 문제이다.

본 논문은 이러한 문제를 해결하기 위해 이벤트 데이터와 RGB 영상을 정합하기 위한 프레임워크를 제안한다. 제안하는 프레임워크는 <그림 1>과 같다. 먼저 이벤트 데이터와 RGB 영상을 이벤트 프레임으로 변환한다. 이벤트 데이터는 누적 과정을 통해 이벤트 프레임으로 생성되며, RGB 영상은 이벤트 카메라의 특성을 모델링하여 RGB 이벤트 프레임으로 변환한다. 이후 두 카메라의 내재적 및 외재적 파라미터를 활용하여 RGB 이벤트 프레임을 이벤트 카메라의 영상 좌표계로 투영한다. 투영된 RGB 이벤트 프레임과 이벤트 프레임에서 특징점 추출 및 매칭을 수행하며, 매칭된 특징점을 기반으로 호모그래피 행렬을 유도한다. 호모그래피 행렬 유도 과정에서 RANSAC(RANdom SAmple Consensus) 알고리즘을 통해 호모그래피 행렬을 반복적으로 계산하고 이상치가 가



<그림 1> 제안하는 프레임워크 бл록도

장 적은 최적의 호모그래피 행렬을 도출한다. 마지막으로, 계산된 호모그래피 변환 행렬을 이용해 RGB 영상을 이벤트 프레임으로 워핑(warping)하여 RGB 영상과 이벤트 프레임 간 정합을 수행한다.

### ■ 이벤트 프레임 생성

본 절에서 이벤트 데이터를 누적하여 이벤트 프레임을 생성하는 방법을 제안한다. RGB 영상 취득 간격과 동일한 시간을 이벤트 누적 시간 간격으로 설정하고, 해당 시간 동안 발생한 모든 이벤트를 하나의 이벤트 프레임으로 누적한다. 이벤트 프레임의 초기값을 비트 심도(bit depth)의 중간값으로 설정하고, 특정 픽셀에서 이벤트가 발생할 때마다 해당 픽셀의 값을 이벤트 극성에 따라  $+\gamma$  또는  $-\gamma$  만큼 누적한다.

### ■ RGB 이벤트 프레임 생성

본 논문에서 제안하는 RGB 이벤트 프레임 변환은 이벤트 카메라의 특성을 모델링하여 RGB 영상으로부터 RGB 이벤트 프레임을 생성한다. RGB 이벤트 프레임을 생성하는 각 단계는 다음과 같다. 먼저, RGB를 그레이스케일(gray scale) 형식으로 변환하여 하나의 밝기 값을 표현한다. 변환 식은 ITU-R BT.601 표준을 따른다. 이후, 그레이스케일로 변환된 밝기 값을 로그 스케일로 변환한다. 이는 이벤트 카메라의 로그 스케일(logarithmic scale)의 밝기 변화를 감지하는 특성을 모델링한다. 로그 매핑 후 1<sup>st</sup> order Infinite Impulse Response(IIR) 저대역 통과 필터를 통해 이벤트 카메라에서 픽셀의 아날로그 대역폭 특성을 모델링한다. 픽셀의 대역폭은 로그 밝기 값에 비례하여 증가하며, 이러한 필터링 과정을 통해 이벤트 픽셀의 실제 응답 특성을 재현한다. 해당 과정은 식 (1)을 통해 수행된다. 식 (1)에서  $L_{filter, current}$ 는 현재 시점에서 필터링 수행된 로그 스케일 밝기 값을 의미하고,  $L_{filter, previous}$ 는 이전 시점에서  $L_{filter, current}$ 와

동일한 방법으로 계산된 로그 스케일 밝기 값을 의미한다.  $\alpha$ 는 픽셀의 선형 밝기 값에 비례하는 가중치이다.  $L_{filter, previous}$ 와  $L_{filter, current}$ 의 차분을 통해 각 픽셀에서 발생한 이벤트의 개수와 극성을 계산한다. 이때, 식 (2)처럼 차분 값을 특정 임계값  $\theta$ 로 나누어 이벤트의 개수  $N$ 을 구한다. 만약  $N > 0$  이면 양의 이벤트,  $N < 0$  이면 음의 이벤트가 발생했음을 의미한다. 마지막으로, 유도된 이벤트 개수를 활용하여 이벤트 프레임 누적 방법과 유사한 방법으로 RGB 이벤트 프레임을 생성한다.

$$L_{filter, current} = \alpha \times L_{log} + (1 - \alpha) \times L_{filter, previous} \quad (1)$$

$$N = \frac{L_{filter, current} - L_{filter, previous}}{\theta} \quad (2)$$

### ■ 특징점 기반 호모그래피 행렬 유도

본 절에서는 특징점 기반 호모그래피 행렬 유도 방법에 대해 다룬다. 본 절에서 RGB 카메라와 이벤트 카메라의 내재적 카메라 파라미터와 두 카메라 사이 외재적 카메라 파라미터가 주어져 있고 깊이 정보는 주어져 있지 않다고 가정한다. 특징점 기반 호모그래피 행렬 유도 방법의 전처리 과정으로 카메라 내재적 파라미터와 외재적 파라미터를 활용해 두 카메라 간 좌표계 변환을 수행한다. 먼저, RGB 카메라의 내재적 파라미터를 활용하여 RGB 이벤트 프레임을 RGB 영상 좌표계에서 정규화된 RGB 카메라 좌표계로 변환한다. 이후 두 카메라 간의 상대적 위치와 방향을 정의하는 카메라 외재적 파라미터를 통해 정규화된 RGB 카메라 좌표계에서 정규화된 이벤트 카메라 좌표계로 변환을 수행한다. 이때, 깊이 정보가 주어지지 않았기 때문에 변환 과정에서 깊이를 1로 가정한다. 마지막으로 이벤트 카메라의 내재적 파라미터를 활용하여 정규화된 이벤트 카메라 좌표계에서 이벤트 카메라 좌표계로 변환한다.

이벤트 카메라 영상 좌표계로 변환된 RGB 이벤트 프

레이임과 이벤트 프레임에 특징점 추출과 특징점 매칭을 수행한다. 이벤트 프레임과 RGB 이벤트 프레임에서 FAST(Features from Accelerated Segment Test) 기반의 공통된 특징점을 추출한다. 기존 FAST 알고리즘과 달리 원형에 주위 픽셀에서 중심 픽셀과 차이 값이 임계 값 을 초과하는 주위 픽셀의 개수가 상한선을 초과할 경우, 해당 픽셀을 노이즈로 판단하여 특징점 검출에서 제외한다. 이후, 검출된 특징점에 대해 BRIEF(Binary Robust Independent Elementary Features) 기반의 특징 이진 기술자를 생성한다. 특징점 매칭 단계에서 이벤트 프레임과 RGB 이벤트 프레임 사이의 서로 대응되는 특징점을 매칭하며, 해밍 거리를 기반으로 최적의 대응점을 찾는다. 전역 탐색을 수행하여 신뢰도 있는 매칭을 수행하며, 특징점 매칭 과정에서 교차 매칭 기법을 적용하여 RGB 이벤트 프레임에서 이벤트 프레임으로 특징점을 매칭했을 때 대응되는 두 특징점과 이벤트 프레임에서 RGB 이벤트 프레임으로 매칭했을 때 대응되는 두 특징점이 서로 일치하는 경우에만 유효한 대응점으로 검출한다. 특징점 매칭의 결과를 통해 두 이벤트 프레임 사이의 변환 관계를

모델링하는 호모그래피 행렬을 유도하며, 해당 과정에서 RANSAC 알고리즘을 활용한다. RANSAC을 통해 반복적으로 호모그래피 행렬을 유도하며, 이상치가 최소화되는 최적의 호모그래피 변환 행렬을 찾는다. 마지막으로, 유도된 변환 행렬을 통해 RGB 영상을 이벤트 프레임으로 위평하여 RGB 영상과 이벤트 프레임 간 정합을 수행한다.

제안하는 프레임워크는 이벤트 카메라와 RGB 카메라로 구성된 자율 주행 스테레오 데이터 셋인 DSEC Dataset을 활용하여 검증되었다. 〈그림 2〉를 통해 서로 다른 깊이에 존재하는 다중 객체가 포함된 복잡한 장면에서 영상 정합이 성공적으로 수행되었음을 주관적으로 확인하였으며, 정합 전 평균 RMSE 57.68에서 정합 후 평균 RMSE 24.46으로 객관적으로 우수한 성능 개선이 입증되었다. 〈그림 2〉에서 (a), (b)는 각각 제안하는 프레임워크의 입력 영상, 출력 영상이며, 두 영상은 이벤트를 누적하여 생성한 이벤트 영상과 RGB 영상을 오버레이한 영상이다. 이벤트 영상에서 이벤트의 극성에 따라 양의 이벤트가 발생한 픽셀은 파란색, 음의 이벤트가 발생한 픽셀은 빨간색으로 표현하였다.



〈그림 2〉 제안하는 프레임워크의 결과

## 주 대 연



- 2023년 2월 : 광운대학교 전자공학과 학사
- 2025년 2월 : 광운대학교 컴퓨터공학과 석사
- 2025년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- 주관심분야 : 비디오 압축, 영상 처리, 이벤트 카메라 신호 처리