

# 강인한 보행자 검출을 위한 계층적 분류기 모델

황순민<sup>0</sup>, 오태현, 권인소

한국과학기술원 전기 및 전자공학과

smhwang@rcv.kaist.ac.kr, thoo@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

## 요약

본 논문에서는 강인한 보행자 검출을 위한 계층적인 분류기 모델을 제안한다. 제안하는 분류기 모델은 기준 검출기와 전문 검출기의 두 단계로 구성되며, 기준 검출기는 전문 검출기가 학습해야 할 데이터들을 크게 줄여줌으로써 전문 검출기가 더 좋은 성능을 갖도록 하는 역할을 한다. 전문 검출기는 기준 검출기에서 보행자라고 판단한 데이터에 대하여 다시 한 번 보행자 여부를 검사함으로써, 잘 못 분류된 데이터를 바로잡아주는 역할을 한다. 기준 검출기에서 검사해야 하는 데이터가 영상 당  $10^7$  개인데 비하여 전문 검출기는 기준 검출기의 검출 결과에 대해서만 적용되므로  $10^2$  개 정도의 데이터에 대해서만 추가적인 계산을 함으로써 기준 검출기의 검출 성능을 향상시킬 수 있다. 변형 가능한 부분 모델(Deformable Part Model)을 기준 검출기로 이용하였으며, 최근 가장 많이 사용되는 Caltech 보행자 데이터 셋에서 제안한 분류기 모델이 로그 평균 미탐지율을 49.38%에서 43.28%로 개선하였다.

## 1. 서론

자동차-보행자 간의 충돌은 도시에서의 주요 교통 사고 중 하나이다. 보행자 교통사고는 대부분 사망사고로 이어지며, 국내에서도 매년 1,000 명 이상의 보행자가 교통사고로 죽어가고 있다. [1] 이러한 사고를 방지하기 위해 보행자 인식 시스템이 운전 보조 시스템의 핵심적인 요소로서 개발되고 있다.

최근 자동차에 탑재되고 있는 보행자인식 시스템은 레이저 또는 적외선과 같은 추가적인 센서를 사용한 것으로서, 카메라에서 획득한 컬러 영상에 추가적인 정보를 제공함으로써 인식률을 높인다. 따라서, 단일 카메라 영상에서의 보행자 인식률을 높이는 것이 더 강인한 보행자 인식 시스템을 위한 근본적인 해결책이다. 뿐만 아니라, 단일 영상에서의 보행자 인식 기술은 로봇이나 CCTV 등에서 사람을 인식하는 기술에 적용될 수 있고, 좀 더 확장함으로써 범용적인 물체 인식 알고리즘으로도 사용될 수 있다는 점에서 원천 기술로서 연구 할 가치가 있다.

최근 컴퓨터 비전 분야에서의 뜨거운 관심에도 불구하고, 아직 단일 카메라 기반의 보행자 인식 알고리즘은 실용화 단계에 이르지 못하고 있다. 상용화를 높은 탐지율 뿐만 아니라 낮은 오탐지율을 갖도록 하는 것이 중요한데, 현재 학계에서 가장 좋은 성능을 보이는 알고리즘들조차 낮은 오탐지율에서는 50%의 탐지율을 넘지 못하고 있는 실정이다.

[2]에 따르면 기존의 제안된 많은 방법론들이 오탐지가 증가할수록 미탐지율이 최소 5%까지 감소한다. 즉, 어느 정도의 오탐지를 허용한다면 이미 95%

이상의 탐지율을 보이는 알고리즘들이 이미 존재한다는 것이다. 따라서, 단일 카메라 기반의 보행자 인식 기술의 상용화를 위해서는 기존 알고리즘들이 오탐지했던 데이터들만 제대로 구별해 내면 된다.

따라서, 본 연구에서는 높은 탐지율을 유지하면서 오탐지율을 낮추기 위한 계층적 분류기 모델을 제시하였다.

## 2. 보행자 검출 문제의 특징

### 2.1 학습 데이터의 불균형

슬라이딩 윈도우와 영상 피라미드를 이용한 보행자 검출 알고리즘들은 영상 한 장에서  $10^7$  개 가량의 학습 데이터를 추출한다<sup>1</sup>. 그러나 일반적으로 한 장의 영상에는 보통 10 명 내외의 보행자가 등장하기 때문에,  $10^1$  개 정도의 데이터만 정확한 보행자 데이터이다. 또한, 보행자 검출은 이진 분류 문제이기 때문에 나머지는 모두 비 보행자 데이터로 고려된다. 즉, 추출된 대부분의 데이터는 음성 데이터이고, 양성 데이터는 상대적으로 굉장히 적게 존재하는 학습 데이터의 불균형이 존재한다. 보행자 검출을 위한 이진 분류기를 학습하는데 있어서 이런 특징은 중요하게 고려되어야 한다.

<sup>1</sup>  $640 \times 480$  영상에서  $64 \times 32$  크기의 슬라이딩 윈도우를 이용하고, 한 옥타브 당 15 개의 스케일을 고려할 때, 영상 한 장 당  $10^7$  개의 데이터가 추출된다.

2.2 학습 데이터의 치우친 분포

보행자 데이터는 사람이라는 공통된 특성을 가진 영상을 기술한 데이터이므로 필연적으로 비슷한 양상을 띠게 된다. 반면에, 그 외의 모든 데이터들이 음성 데이터로 고려되기 때문에 음성 데이터들은 각기 다른 양상을 띠게 된다. 다시 말하면, 기술된 데이터들을 고차원 공간에서의 한 점으로 생각할 때 양성 데이터들은 대체로 한 쪽에 군집화된 분포를 가지게 되고 음성 데이터들은 모든 공간에 분포하게 된다. 따라서, 양성 데이터를 잘 분류해내기 위해서는 양성 데이터와 비슷하지만 음성 데이터인 어려운 음성 데이터들이 중요하게 고려되어야 한다.

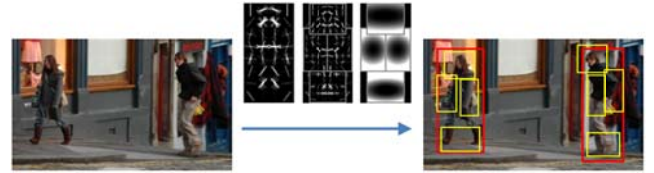


그림 2. DPM은 파트 기반의 모델이기 때문에 부분적인 가려짐을 효과적으로 다룰 수 있다.

3. 계층적 분류기 모델

제안하는 보행자 인식 시스템은 그림 1 과 같이 기존 보행자 검출기를 이용하여 보행자로 분류된 데이터들을 전문 보행자 검출기를 이용하여 다시 한 번 분류를 시도한다. 즉, 두 검출기의 검출결과를 함께 고려함으로써, 기존 검출기가 오탐지한 데이터들을 바로잡을 수 있다.

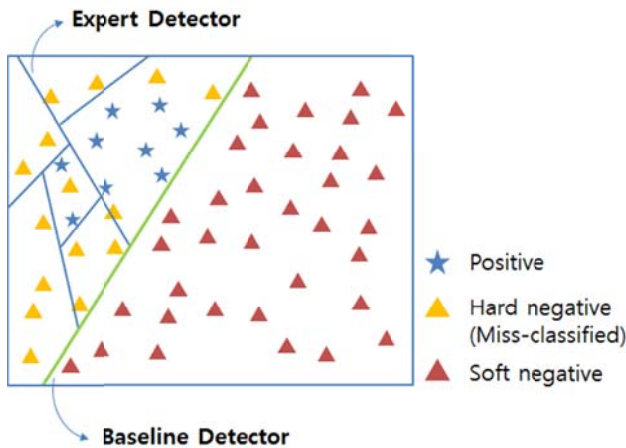


그림 1. 계층적 분류기 모델. 전문 검출기는 기존 검출기가 보행자로 분류한 데이터에 의해서만 학습 및 분류를 수행한다.

3.1 기존 검출기: Deformable Part Models

본 연구에서는 기존 보행자 검출기로서 DPM [3]을 이용하였다. 이 방법론은 부분적인 가려짐을 효과적으로 다룰 수 있는 파트 기반 모델로서, 최근 물체 인식 분야의 대약진을 이루어 낸 알고리즘이다.

기존 검출기는 양성 데이터와 확연하게 다른 음성 데이터를 일차적으로 걸러주는 역할을 한다. 즉, 전문 검출기의 학습 영역을 줄여주는 것이 기존 검출기의 핵심적인 역할이다.

3.2 전문 검출기: Random Forest

전문 검출기로는 Random Forest [4]를 사용하였다. 기존 검출기에서는 잘못 분류된 데이터들을 다시 분류해야 하기 때문에 복잡한 결정 평면을 만들 수 있는 모델이 사용되어야 한다. 본 연구에서 사용한 Random Forest 분류기는 계층적으로 약한 모델에 의해서 데이터를 분류해내는 결정 나무들의 집합으로 정의된다. 즉, 각각의 결정 나무에 의한 분류 점수를 모두 더하거나 곱함으로써 최종 분류 점수가 결정되는데, 그림 3 과 같이 다양한 결정 나무들에 의한 결정 평면들의 조합으로써 최종 결정 평면이 결정되기 때문에 복잡한 결정 평면을 표현 할 수 있다.

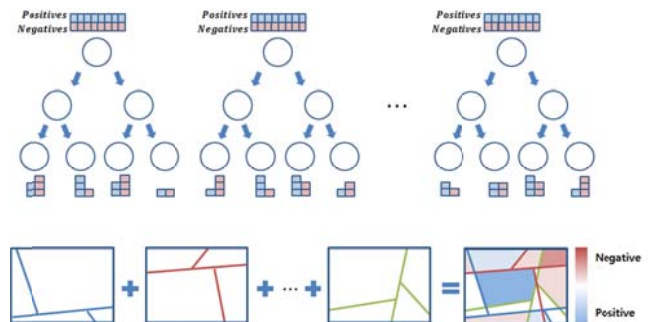


그림 3. Random Forest는 결정 나무들의 앙상블 모델이기 때문에, 복잡한 결정 평면을 표현 할 수 있다.

3.3 계층적 분류기에 의한 보행자의 검출

전문 검출기에서는 기존 검출기에서 사용했던 특징량 벡터를 그대로 이용한다. 즉, 기존 검출기는 전문 검출기에서 검사해야 하는 데이터를 줄여주는 역할을 할 뿐, 두 검출기에서 모두 검사된 검출 결과는 서로 독립적이다. 따라서, 그림 4 와 같이 두 검출 결과는 서로 곱함으로써 최종 검출 점수를 계산할 수 있다.

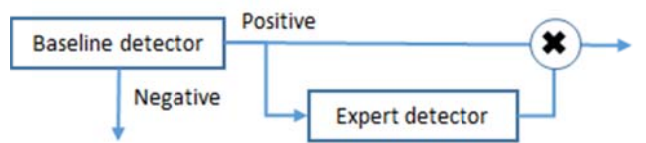


그림 4. 기존 검출기에서 보행자로 분류된 데이터들의 최종 검출 점수는 두 검출기의 검출 점수를 곱함으로써 계산된다.

4. 실험 결과 및 분석

검출기의 성능을 평가하기 위하여, 최근 가장 많이 이용되고 있는 Caltech 보행자 데이터 셋을 이용하였다. 평가에 사용된 척도는 로그 평균 미탐지율 (Log-average miss-rate)과 영상 당 오탐지율에 대한 미탐지율 그래프이다. 평균 미탐지율은 [2]에서와 같이  $[10^{-2}, 10^0]$  구간에서 균일하게 샘플링 된 미탐지율의 로그 평균이다.

그림 5 에서 보여지듯이, Caltech 테스트 셋에서 가장 대표적으로 사용되는 ‘Reasonable’ 평가 집합에서 기준 검출기의 미탐지율을 6.1% 가량 개선하였다. 뿐만 아니라, 다양한 평가집합에서 제안한 계층 구조를 적용했을 때, 오탐지율이 크게 감소했다.

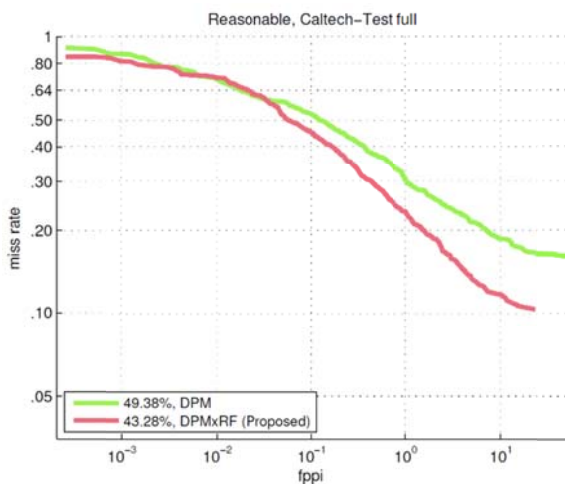


그림 5. Caltech 테스트 셋 - ‘Reasonable’

	Baseline	Proposed	Note
Reasonable	49.38%	<b>43.28%</b>	$\geq 50\text{pix}$ , $\text{occ} \leq 35\%$
Near	29.83%	<b>25.37%</b>	$\geq 80\text{pix}$ , No occ
Medium	75.33%	<b>71.47%</b>	$30\text{pix} \leq \text{height} \leq 80\text{pix}$ , No occ
No occlusion	47.43%	<b>41.52%</b>	$\geq 50\text{pix}$ , No occ
Heavy occlusion	91.97%	<b>85.54%</b>	$\geq 50\text{pix}$ , $35\% \leq \text{occ} \leq 80\%$

표 1. 다양한 평가집합에서의 성능 개선

5. 결론

본 연구에서는 기준 검출기를 이용하여 전문 검출기의 학습 공간을 크게 줄임으로써, 구별하기 어려운 비 보행자 데이터를 보행자 데이터와 구별 해 낼 수 있는 비선형 전문 검출기를 학습하는 방법을 제안하였다. 제안한 구조는 더 구별성이 좋은 전문 검출기를 학습하는 일반적인 알고리즘으로서, 다른 기준 검출기에 적용해도 성능을 더 개선 할 수 있을 것으로 기대된다. 따라서, 기존에 제안된 최고 성능의 검출기에 적용함으로써 가장 좋은 성능을 달성 할 수 있을 것이다.

감사의 글

본 논문은 지식경제부 산업융합원천기술개발사업으로 지원된 연구결과입니다[10044775, 보행자 보호를 위한 자동 긴급 제동(AEB) 시스템 원천 기술 개발].

참고문헌

- [1] Traffic Accident Analysis System, <http://taas.koroad.or.kr>.
- [2] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 743-761, 2012.
- [3] P. F. Felzenszwalb, R. B. Girchick, D. McAllester, and D. Ramanan, “Object Detection with Discriminatively Trained Part Based Models,” IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, 2010.
- [4] A. Criminisi, J. Shotton, and E. Konukoglu, “Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning,” Foundations and Trends in Computer Graphics and Vision, vol 7, no 2-3, pp. 81-227, 2012.