

Integral Channel Feature 를 이용한 보행자 검출 구현

김동영*[†], 이충희*

*대구경북과학기술원 IT 융합연구부

([†]) 교신저자

e-mail : dckyoung@dgist.ac.kr, chlee@dgist.ac.kr

Implementation of Pedestrian Detection using Integral Channel Feature

Dongyoung Kim*[†], Chung-Hee Lee*

*Division of IT Convergence, DGIST

([†]) Corresponding Author

요 약

최근 여러 매체에서 화두가 되고 있는 자율 주행 자동차나 Advanced driver assistance systems (ADAS)과 같은 분야에서 보행자 검출 기술은 핵심 요소 기술 중에 하나로 손꼽히고 있다. 특히, 인간의 인지 부하(Cognitive Load)를 고려했을 때, 주행 중에 발생할 수 있는 모든 사건을 다룬다는 것은 매우 어렵기 때문에, 앞서 언급한 방법의 도움을 받아 도로 주행 중에 발생 될 수 있는 인명 사고율을 줄이고자 하는데 그 목적이 있다. 본 논문에서는 Integral Channel Feature 를 사용하여 AdaBoost 알고리즘으로 보행자 검출을 위한 분류기를 구현하였다. 그 결과, INRIA 에서 제공되는 Pedestrian dataset 에서 Detection rate 는 97%이상, False positive 는 1%에 정도로 나타났다.

1. 서론

최근 자율 주행 차량이나 무인 자동차와 같은 키워드가 인기 물이를 하면서, 이와 관련된 기술의 요구가 급증하고 있다. 보행자 검출은 차량 안전이나 보안, 로봇 등과 같이 다양한 분야에서 핵심 기술 중에 하나로 손 꼽히고 있으며, 많은 비용과 인력이 투입되어 현재 활발히 연구가 진행되고 있는 분야이다. 뿐만 아니라, INRIA 나 Caltech, Daimler 의 보행자 검출 관련 학습 및 테스트 데이터가 공개되어 있어서, 누구나가 쉽게 접근하여 사용할 수 있다. 이 중에서 INRIA 는 가장 많이 사용되는 데이터 베이스로서, 본 논문에서 보행자 검출을 위해 학습 및 테스트 데이터 베이스로 사용한다[1].

본 논문의 2 장에서는 속도 향상을 위한 방법으로 Integral Image 에 대해서 먼저 살펴보고, Integral Channel Feature 를 이용하여 특징을 구성한 다음에, AdaBoost 알고리즘을 통해 분류기를 학습한다. 또한 3 장에서는 이와 관련한 실험 결과를 INRIA 데이터 셋에 기반하여 도출하고, 마지막으로 4 장에서는 결론 및 향후 과제에 대해 이야기 한다.

2. 본론

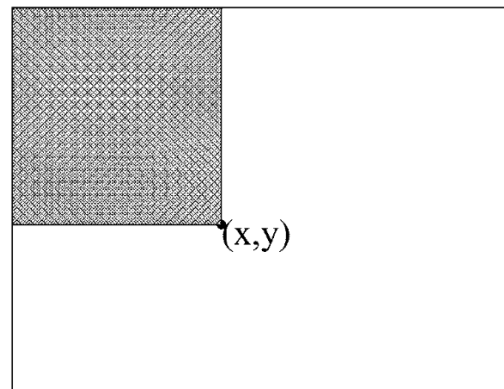
2.1 Integral Image

Integral Image 는 영상 내에서 특정 사각 영역의 픽셀 값을 빠르게 계산 할 수 있는 방법으로

Viola & Jones 에 의해 소개되었다[2]. (수식 1)과 (그림 1)에서 알 수 있듯이, 해당 픽셀의 적분 영상의 값은 사각형 영역 내에 존재하는 모든 픽셀의 합 의 값을 가진다.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

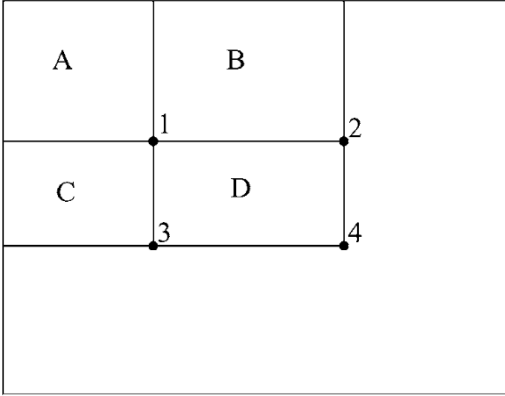
(수식 1) ii 는 Integral Image 를 i 는 원본 이미지를 각각 나타내고, (x, y)는 해당 픽셀의 위치를 나타낸다.



(그림 1) Integral Image

예를 들어, (그림 2)의 D 영역에 해당하는 픽셀 값

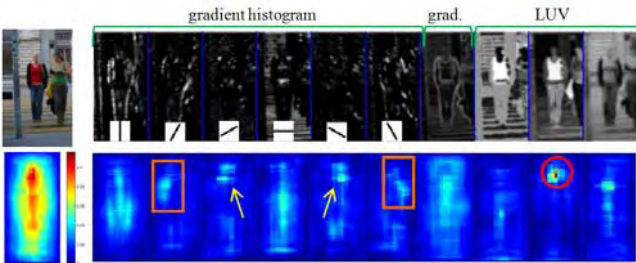
을 계산을 한다고 하면, 일반적인 영상에서는 가로와 세로에 각각 반복문을 사용하여 $O(n^2)$ 의 시간 복잡도를 가지겠지만, Integral Image 를 사용하게 되면 (그림 2)의 [1,2,3,4] 위치에 해당하는 값을 사용하여, $O(1)$ 의 상수 시간으로의 계산이 가능하여, 속도 향상에 크게 기여한다는 장점이 있다.



(그림 2) 영역 D = (1 + 4) - (2 + 3)

2.2 Integral Channel Feature

기존의 보행자 검출 방법에 있어서 Histograms of Oriented Gradients(HOG)와 Support Vector Machine(SVM) classifier 의 조합이 가장 많이 사용되었다[3]. Integral Channel Feature(ICF)는 HOG 의 개념을 확장하여 이미지의 각도 정보와 색 정보를 각각 채널로 구성하여, AdaBoost 알고리즘을 통한 학습이 소개되었다[4].



(그림 3) Integral Channel Feature 의 구성

Integral Channel 은 6 개의 Bin 을 가지는 gradient histogram 과 gradient magnitude, 그리고 L, U, V 로 총 10 개의 채널을 가진다. 각 채널에서의 Feature 값은 First-order feature 라고 소개된 Local sum 으로 해당 영역의 픽셀의 합을 사용한다.

2.3 AdaBoost

AdaBoost 알고리즘은 아직까지도 널리 사용되는 알고리즘으로, 기본적인 개념은 여러 개의 약한 분류기를 조합하여 하나의 강한 분류기를 만드는 것이다.

<표 1>은 AdaBoost 알고리즘을 나타내고 있는데, 매번 주어진 학습 데이터 셋에서 negative 와 positive 를 구분할 때 가장 에러가 낮은 값을 주는 약 분류기를 선택하고 이를 이용하여 가중치를 업데이트한다.

이 과정을 T 만큼 반복하여 최종적으로 하나의 강 분류기를 도출해내도록 하는 방법이다.

<표 1> AdaBoost 알고리즘

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \dots, T$:

1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$
2. Select the best weak classifier with respect to the weighted error

$$\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|.$$

See Section 3.1 for a discussion of an efficient implementation.

3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where $f_t, p_t,$ and θ_t are the minimizers of ϵ_t .
4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- The final strong classifier is:

$$C(x) = \begin{cases} 1 & \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

3. 실험 결과

3.1 Dataset

INIRA Pedestrian dataset 을 사용하였다. 해당 데이터 셋의 학습 셋은 negative 1218 장과 positive 614 장으로, 테스트 셋은 negative 453 장과 positive 288 장으로 구성되어 있다.

학습은 positive 영상에서 보행자가 존재하는 모든 영상을 추출하여 미러링 시켜 총 2416 장을 준비했고, negative 의 경우 임의로 positive 영상의 두 배수에 해당하는 영상을 준비하였다. 테스트 영상 또한 학습 영상과 유사한 방법으로 준비하였으며, positive 와 negative 영상을 각각 1126 장과 29128 장을 준비하였다.

3.2 실험 환경

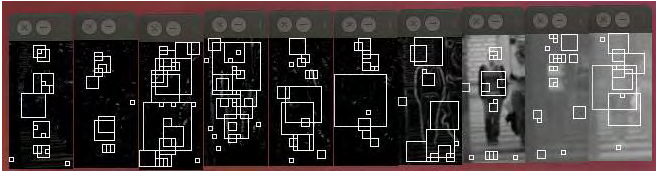
실험은 Intel(R) Core(TM) i7-4970K CPU @ 4.00GHz 와 32GB 메모리의 사양과 Ubuntu 14.04 환경에서 수행되었다.

3.3 실험 결과

총 92 개의 약 분류기로 구성된 강 분류기를 만들었으며, (그림 4)는 본 연구의 실험 결과를 한번에 보여주고 있다. 대체적으로 각 채널 별로 골고루 특

정을 찾아내고 있는 것을 알 수 있다.

총 학습 시간은 약 10 분 가량 소요되었으며, 최종 검출기의 성능은 Detection rate 97.069%와 False positive rate 1.794%를 가진다.



(그림 4) 실험 결과

4. 결론 및 향후 과제

본 논문에서는 INRIA 에서 제공하는 Pedestrian dataset 을 기반으로, Integral Channel Feature 을 사용한 AdaBoost 학습 방법에 대해서 알아보았다. 실험 결과에서 보여지듯이, 짧은 시간에 95%라는 만족스러운 성능을 보여주고 있으며, 향후에 이 알고리즘에 Soft Cascade Structure 를 적용한다면, 검출기의 성능을 대폭적으로 향상시킬 수 있을 것으로 기대된다. 뿐만 아니라, GPU 를 적극적으로 활용하여 더 많은 데이터를 학습할 수 있도록 한다면, 당연히 더 좋은 결과 또한 기대를 할 수 있을 것으로 사료된다.

5. 감사의 글

본 연구는 미래창조과학부에서 지원하는 대구경북과학기술원 기관고유사업에 의하여 수행되었습니다. (15-IT-01).

참고문헌

- [1] Dollár, Piotr, et al. "Pedestrian detection: A benchmark." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- [2] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." *International journal of computer vision* 57.2 (2004): 137-154.
- [3] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, 2005.