**Informed Haar-like Features Improve Pedestrian Detection**

Haar-like 를 이용한, 영상 보행자검출개선 기술의 정보

Shanshan Zhang∗, Christian Bauckhage†‡, Armin B. Cremers∗‡   
∗University of Bonn, Germany   
† Fraunhofer IAIS, Germany   
‡ Bonn-Aachen International Center for Information Technology (B-IT) {zhangs,abc}@iai.uni-bonn.de, [christian.bauckhage@iais.fraunhofer.de](mailto:christian.bauckhage@iais.fraunhofer.de)

**Abstract**

We propose a simple yet effective detector for pedestrian detection.   
우리는 보행자 감지를위한 간단하면서도 효과적인 탐지기를 제안합니다.  
The basic idea is to incorporate common sense and everyday knowledge into the design of simple and computationally efficient features.

기본 개념은 상식과 일상 지식을 간단하고 계산상 효율적인 기능을 설계하는 데 포함시키는 것이다.

As pedestrians usually appear up-right in image or video data, the problem of pedestrian detection is considerably simpler than general purpose people detection.

보행자는 일반적으로 이미지 또는 비디오 데이터에서 오른쪽 위의 사진처럼 보이기 때문에 보행자 감지의 문제는 범용 사용자 검색보다 훨씬 간단하다.

We therefore employ a statistical model of the up-right human body where the head, the upper body, and the lower body are treated as three distinct components.

그러므로 우리는 머리, 상체 및 하체가 세 가지 별개의 구성 요소로 취급되는 상류 인체의 통계 모델을 사용한다.

Our main contribution is to systematically design a pool of rectangular templates that are tailored to this shape model.

우리의 주요 공헌은 이 모양 모델에 맞춘 직사각형 템플릿 풀을 체계적으로 설계하는 것이다.

As we incorporate different kinds of low-level measurements, the resulting multi-modal & multi-channel Haar-like features represent characteristic differences between parts of the human body yet are robust against variations in clothing or environmental settings.

다양한 종류의 저수준 측정을 통합함에 따라 다중 모드 및 다중 채널 Haar와 유사한 기능은 인체의 각 부분 사이의 특징적인 차이를 나타내지만 의류(의 색상?) 또는 환경 설정의 변동에 강하다.

Our approach avoids exhaustive searches over all possible configurations of rectangle features and neither relies on random sampling.

우리의 접근 방식은 가능한 모든 사각형 구성의 구성에 대한 철저한 검색을 피하고 무작위 샘플링을 사용하지 않는다.

It thus marks a middle ground among recently published techniques and yields efficient low-dimensional yet highly discriminative features.

따라서 최근에 발표 된 기술 가운데 중도에 이르렀으며 효율적이고 저 차원이지만 고도로 차별화 된 기능을 제공한다.

Experimental results on the INRIA and Caltech pedestrian datasets show that our detector reaches state-of-the-art performance at low computational costs and that our features are robust against occlusions.

INRIA 및 Caltech 보행자 데이터 세트의 실험 결과는 우리의 탐지기가 낮은 계산 비용으로 최첨단 성능에 도달하고 우리의 특징이 **오클루젼**에 강하다는 것을 보여준다.

\*\*오클루젼 : 관심영역 이외의 객체를 치워버리는(없애버리는?) 기법

|  |
| --- |
| Figure 1: Illustration of our template pool. Templates are generated by sliding rectangular windows of pre-defined sizes over a pre-defined pedestrian shape model. Note that some templates are ternary (shown as white, black, and red areas) which are given the weights of +1, −1, and 0, respectively.  그림 1 : 템플릿 풀의 그림. 템플릿은 미리 정의된 보행자 모양의 모델을 통해 미리 정의된 크기의 직사각형 창을 밀어넣어 생성된다. 일부 템플릿은 각각 +1, -1 및 0의 가중치가 부여되는 삼원 문자 이다. (흰색, 검은 색 및 빨간색 영역으로 표시됨) |

**1. Introduction**

Over the last decade, the question of how to detect pedestrians in images has been thoroughly investigated [10].

지난 10 년 동안, 이미지에서 보행자를 탐지하는 방법에 대한 문제가 철저히 조사되었다.

Yet, primarily because of random influences such as scene structure, lighting or people’s choice of clothing, the problem remains challenging and continues to attract research.

그러나 주로 장면 구조, 조명 또는 사람들의 옷 선택과 같은 임의의 영향 때문에 문제는 여전히 도전적이며 연구를 계속 끌어들인다.

A noticeable trend in this domain is that researchers increasingly rely on huge feature pools and high dimensional feature vectors [27] since it is commonly believed that more features integrate more information and thus lead to better performances.

이 영역에서 눈에 띄는 추세는 더 많은 기능이 더 많은 정보를 통합하고 따라서 더 나은 성과로 이어질 것으로 일반적으로 믿어지기 때문에 연구자들이 [huge feature pool] 과 [high dimensional feature vector(고차원 피쳐벡터)] 에 점점 더 의존한다는 것입니다.

\*\* 피쳐풀링(feature pool) : 비디오 프레임들에서 피쳐풀링하면 동적인 피쳐가 얻어진다.

\*\* 개체의 속성은 피쳐값 이라고 부르기도 하며, 보통 피쳐벡터(feature vector) 라는 벡터의 형태로 제공된다. 피쳐벡터는, 선형회귀 에서 사용되어진 explanatory variables(독립변수)의 벡터와 같다

As a consequence, many recent approachesrely on the availability of powerful computers and GPU computation in order to be capable of real-time detection.

결과적으로 실시간 탐지가 가능하도록 강력한 컴퓨터 및 GPU 계산 기능이 제공됩니다.

Also, aspects due to the peculiar geometry of high dimensional spaces, e.g. concentration of measure and neighborliness, appear to be disregarded.

또한, high dimensional space(고차원 공간)의 특유한 기하학적 형상, 예를 들면, [concentration of measure(측정값이 집중된.. 대략 중앙값)] 과 [neighborliness(얻은 값의 근처 값)]은 무시된다.

\*\* 제대로 이해하지 못했다 (<https://en.wikipedia.org/wiki/Concentration_of_measure> )

This raises the question, if there are alternative approaches which require less memory and less computational resources but still perform robust and reliable?

위와 같은 내용들 때문에, 만약 메모리가 적게 들고 계산 자원이 적지만 견고하고 신뢰할 수있는 대체 접근법이 있다면? 이라는 질문을 제기하게 된다.

In this paper, we propose more compact features which simultaneously ensure effectiveness and efficiency.

이 논문에서는 효율성과 효율성을 동시에 보장하는 보다 컴팩트한 기능을 제안한다.

In particular, we argue that by incorporating prior information as to the appearance of the up-right human body, one can design reasonable features for pedestrian detection.

특히, 우리는 [up-right human body]의 외관에 대한 사전 정보를 통합함으로써 보행자 감지를 위한 합리적인 기능을 설계할 수 있다고 주장한다.

In fact, from the point of view of visual perception, pedestrians form a class of high intra-class similarity.

실제로, 시각적 인식의 관점에서, 보행자는 높은 클래스내 유사성의 클래스를 형성합니다.

This is because strong regularities of up-right body shapes limit how pedestrians may appear in image data.

이것은 강한 규칙성 때문이다. [up-right body] shape 형태의 보행자가 이미지 데이터에 나타날 수 있는 방법이 제한되기 때문이다.

In particular the headshoulder area of the human body shows a geometry seldom found among other natural objects.

특히 인체의 머리 부분 영역은 다른 자연물 중에서 거의 발견되지 않는 기하학을 보여줍니다.

Based on a careful exploration of these characteristics, we design new features that enable efficient, state-of-the-art pedestrian detection.

이러한 특성을 면밀히 조사하여 능률적인 최첨단 보행자 감지를 가능하게 하는 새로운 기능을 설계한다.

Our approach is motivated by prior work on detecting objects of rather low intra-class variability.

우리의 접근 방식은 다소 낮은 클래스내 변동성의 객체를 탐지하는 사전 작업에 의해 동기 부여된다.

In particular, HOGs [5] and cascaded Haar-like features [24] have become the de-facto methods of choice in this area.

특히, HOGs와 cascaded Haar-like features [24]은 이 분야에서 사실상 선택 방법이 되었습니다.

Yet, we note that corresponding features are either determined by means of exhaustive searches over all possible variations [24] or by means of less exhaustive random sampling [13].

그러나, 우리는 상응하는 특징들이 **가능한 모든 변이에 대한 철저한 검색**이나 **덜 철저한 무작위 표본 추출** [13]을 통해 결정된다는 점에 주목한다.

In this paper, we propose a method that marks a middle ground;

이 논문에서는 중간 지점을 표시하는 방법을 제안한다.

we design compact, discriminative Haar-like features selected from a particular template pool that re-flects prior information about the pedestrian up-right body shape.

우리는 보행자의 [up-right body shape]에 대한 이전 정보를 다시 반영하는 특정 템플릿 풀에서 선택된, 차별적인 Haar-like features을 디자인한다.

Extensive experiments indicate that these features are highly characteristic and therefore enable very robust detection.

광범위한 실험 결과에 따르면 이러한 특징들은 매우 특징적이며 따라서 매우 견고한 탐지를 가능하게한다.

**1.1. Related work**

**관련된 일**

Because of its practical impact, research on pedestrian detection has noticeably intensified over the past decade and the literature on possible solutions is vast. Since an exhaustive survey is beyond the scope of this paper, our following review therefore focuses on features that have been proposed in this context.

실용적인 영향 때문에 지난 10 년 동안 보행자 탐지에 대한 연구가 눈에 띄게 강화되었으며 가능한 해결책에 대한 문헌이 광범위합니다. 철저한 조사가 본 백서의 범위를 벗어나는 것이므로 다음의 검토는 이 맥락에서 제안된 기능에 중점을 둡니다.

As of this writing, the arguably most popular features for visual pedestrian detection are based on Histograms of Oriented Gradients (HOGs) as introduced in [5]. HOG features brought about significant improvements and therefore establish an important baseline. Felzenszwalb et al. [LatSvm] [12, 11] successfully employed HOG features in a part-based model for object detection; Wang et al. [HogLbp] [26] combined HOG features with a particular Local Binary Pattern (LBP) feature in order to cope with partial occlusions. Walk et al. [25] combined HOG features with self-similarity features related to color channels [MultiFtr+CSS] as well as motion features [MultiFtr+Motion] in order to better integrate spatial and temporal information.

이 글을 쓰는 시점에서 시각적인 보행자 감지를 위한 논란의 여지는 있지만 가장 인기있는 특징은 [5]에서 소개 된 HOG(Histogram of Oriented Gradients)를 기반으로합니다. HOG 기능은 중요한 개선을 가져왔고 따라서 중요한 기준을 설정합니다. Felzenszwalb et al. [LatSvm] [12, 11]은 객체 탐지를 위한 부분 기반 모델에서 HOG 기능을 성공적으로 사용했습니다. Wang et al. [HogLbp] 부분 폐색에 대처하기 위해 HOG 기능을 특정 Local Binary Pattern (LBP) 기능과 결합했습니다. Walk et al. [25] 컬러 채널 [MultiFtr + CSS]뿐만 아니라 모션 특징 [MultiFtr + Motion]과 관련된 자기 유사성 특징과 HOG 기능을 결합하여 공간 및 시간 정보를 보다 잘 통합합니다.

Deviating from the popular framework of “HOG+SVM” computations, Dollar ́ et al. [8] applied integral channel features which efficiently integrate multiple cues due to colors and gradients by means of employing integral images.

"HOG + SVM"계산의 대중적인 틀에서 벗어나, Dollar et al. [8]은 적분 영상을 사용함으로써 색과 기울기에 기인 한 다중 신호를 효율적으로 통합하는 적분 채널 특징을 적용했다.

For classification, they used boosting methods and thus obtained a real-time detector [ChnFtrs].

분류를 위해, 그들은 부스팅 방법을 사용하여 실시간 검출기 [ChnFtrs]를 얻었다.

An extension of this approach has been called the “Fastest Pedestrian Detection in the West” [FPDW] [7] and was shown to en-able particularly fast multiscale detection. Due to its efficiency and reasonable performance, many new detectors [3, 6] therefore consider [ChnFtrs] as a baseline and several authors obtained even better performance by extending the feature pool in various ways. Benenson et al. [Roerei] [4] used irregular rectangles resulting in a 718, 080 dimensional feature pool; Lim et al. [SketchTokens] [15] added self-similarity features, yielding a 21, 350 dimensional feature vector for image patches of a size of 35 × 35 pixels.

이 접근법의 확장은 "서쪽에서 가장 빠른 보행자 탐지 (FPDW) [7]"라고 불려 왔고 특별히 빠른 멀티 스케일 탐지를 가능하게하는 것으로 나타났습니다. 그것의 효율성과 합리적인 성능으로 인해 많은 새로운 탐지기 [3, 6]는 기준선으로 [ChnFtrs]를 고려하고 여러 저자는 다양한 방법으로 feature pool 을 확장하여 더 나은 성능을 얻었습니다. Benenson et al. [Roerei] [4]는 불규칙한 직사각형을 사용하여 718, 080 차원 기능 풀을 만들었습니다. Lim et al. [SketchTokens] [15]는 35 × 35 픽셀 크기의 이미지 패치에 대해 21, 350 차원의 특징 벡터를 산출하는 자기 유사성 피쳐(self-similarity features)를 추가했습니다.

Due to the extreme sizes of these feature pools, both corresponding detectors require powerful computing hardware and large amounts of memory at training time. Addressing issues like these, our work aims at building new detectors based on small but intelligently designed feature pools that enable state-of-the-art detection accuracy.

이러한 feature pool이 극단적으로 크기 때문에 해당 감지기는 모두 강력한 컴퓨팅 하드웨어와 많은 양의 메모리가 필요합니다. 이와 같은 문제를 해결하기 위해 우리는 작지만 지능적으로 설계된 feature pools 을 기반으로 최신 감지기 정확도를 기반으로 새로운 탐지기를 구축하는 것을 목표로합니다.

Pioneering attempts of using Haar wavelets for pedestrian detection are found in [17] where it was demonstrated that wavelet templates can be used to define the shape of an object. Later, Papageorgiou et al. [20] proposed a similar yet more general system for object detection and, subsequently, Haar-like features became popular in the object detection community. The epitome of such approaches is found in the work by Viola and Jones [24] who used Haar-

like features in combination with boosting algorithms to build a successful face detector. Dollar ́ et al. [9] proposed to use feature mining strategies to select informative features from a large amount of Haar wavelets. In this context, we note that, in the recent literature, Haar-like fea-

tures are also referred to as second-order channel features [ChnFtrs]. However, Haar-like features are often discarded in pedestrian detection as they seem not to improve perfor-

mance when combined with first-order channel features. In a closer analysis as to possible reasons for this behavior, we found that Haar-like templates that perform well for face detection are not necessarily suited for pedestrian detection as they may fail to capture visual characteristics of human body. As a remedy, we propose to design particularly tailored templates for up-right body shapes.

보행자 검출을 위해 Haar wavelets 을 사용하려는 선구적인 시도는 wavelet 템플릿이 객체의 모양을 정의하는데 사용될수 있다는 것이 입증된 [17]에서 발견됩니다. Later, Papageorgiou et al. [20]은 물체 탐지를 위한 비슷하지만 보다 일반적인 시스템을 제안하였고, 이후에 Haar-like 특징들이 물체 탐지 공동체에서 대중화되었다. 이러한 접근법의 전형은 Viola와 Jones [24]가 Haar와 유사한 기능을 사용하여 성공적인 얼굴 검출기를 구성하는 알고리즘을 향상시킨 연구에서 발견됩니다. Dollar et al. [9]는 많은 양의 Haar 웨이브 렛에서 유익한 특징을 선택하기 위해 feature 마이닝 전략을 사용할 것을 제안했다. 이 문맥에서 우리는 최근 문헌에서 Haar와 같은 특징을 2 차 채널 특징 [ChnFtrs]이라고도 부릅니다. 그러나 Haar와 같은 기능은 보행자 감지 기능에서 자주 삭제되지만, 일차 채널 기능과 결합 할 때 이 행동에 대한 가능한 이유에 대해 더 자세히 분석해 보면 얼굴 검출에 좋은 Haar-like 템플릿이 인체의 시각적 특성을 포착하지 못할 수 있으므로 보행자 감지에 반드시 적합한 것은 아니라는 사실을 발견했습니다. 해결 방법으로, 우리는 up-right body shapes 를 위해 특별히 맞춤형 템플릿을 디자인 할 것을 제안합니다.

**1.2. Contributions**

Our main contribution in this paper is to model pedestrian shapes in terms of three rectangles that are geared towards different body parts. Based on this shape model, we design compact Haar-like features to describe local differences. Accordingly, we design a compact feature pool that is better tailored to pedestrian shapes than the ones covered in the above survey.

이 논문에서 우리의 주요 공헌은 서로 다른 신체 부위를 대상으로 하는 3 개의 사각형으로 보행자 모양을 모델링하는 것입니다. 이 모양 모델을 기반으로 우리는 지역 차이를 설명하기 위해 소형 Haar와 유사한 기능을 설계합니다. 따라서 우리는 위의 조사에서 다루는 것보다 보행자 모양에 더 잘 맞은 소형 기능 풀을 설계합니다.

Template pool for pedestrian shape model: we find that up-right walking pedestrians share a common visual appearance especially with respect to the geometry of the head and shoulder region of the body. Based on this shape model, we design a pool of rectangle features (rectangular templates) that is adapted to these local structures. Our templates are specific for pedestrians and therefore lead to a very good performance; on the other hand, they constitute only a small subset of the set of all possible rectangular templates so they significantly reduce training times.

보행자 모양 모델을 위한 템플릿 풀 : 우리는 up-right walking 보행자가 신체의 머리와 어깨 영역의 기하학과 관련하여 공통적인 시각적 외관을 공유한다는 것을 알게되었습니다. 이 모양 모델을 기반으로 이 로컬 구조에 맞게 a pool of rectangle features (사각형 템플릿)을 설계합니다. 우리의 템플릿은 보행자 전용이므로 매우 우수한 성능을 제공합니다. 다른 한편으로는 가능한 모든 직사각형 템플릿 세트의 작은 부분 집합을 구성하므로 training time이 크게 단축됩니다.

Multi-modal & multi-channel Haar-like features: we use two template modalities –binary and ternary– for Haar-like features. The ternary modal is specifically proposed to represent corner regions found along the pedestrian silhouette so as to enable rectangle features to represent more complex geometric configurations. In terms of channel features, we consider rectangle descriptors not only with respect to colors but also with respect to gradients. This addresses challenges due to variations in the choice of clothes.

다중 모달 및 다중 채널 Haar와 유사한 기능 : 우리는 Haar와 같은 기능에 대해 두 가지 템플릿 모달리티 -이진 및 삼항- 을 사용합니다. 삼항 모달은 보행자 실루엣을 따라 발견된 코너 영역을 나타내도록 제안되어 사각형 feature가 보다 복잡한 기하학적 구성을 나타낼 수 있게 합니다.

채널 특성의 관점에서 우리는 color에 대해서 뿐만 아니라 그라디언트에 대해서도 rectangle descriptor를 고려합니다. 이는 옷 선택의 다양성으로 인한 어려움을 해결합니다. (실제, 옷 색깔에 따라 잘못인식되는 경우가 다소 있다고 함)

We evaluate our approach in extensive experiments on several benchmark datasets and demonstrate that by employing compact features, our new pedestrian detector achieves state-of-the-art performance while enjoying three advantages: it is easy to implement, easy to train, and fast to apply on real world data.

우리는 여러 벤치 마크 데이터 세트에 대한 광범위한 실험을 통해 우리의 접근법을 평가하고 컴팩트 한 기능을 사용하여 새로운 보행자 감지기가 최첨단 성능을 달성하면서도 구현하기 쉽고, 훈련하기 쉽고, 빠른 속도로 세 가지 이점을 누릴 수 있음을 보여줍니다. 실제 데이터에 적용하십시오.

**2. Template pool**

In this section, we describe how to generate a template pool that is tailored towards visual pedestrian detection. For this purpose, we first define a pedestrian body shape model and then generate templates by sliding bounding boxes of different sizes over this shape model. Fig. 1 illustrates the whole template pool and shows examples of templates of different sizes.

이 섹션에서는 시각적 보행자 감지에 맞춰진 템플릿 풀을 생성하는 방법에 대해 설명합니다. 이를 위해 먼저 보행자 몸체 모델을 정의한 다음이 모양 모델에 대해 서로 다른 크기의 경계 상자를 슬라이딩하여 템플릿을 생성합니다. 무화과. 1은 전체 템플릿 풀을 보여주고 크기가 다른 템플릿의 예를 보여줍니다.

**2.1. A pedestrian body shape model**

**보행자 body shape 모델**

We define a pedestrian body shape based on statistical information. The INRIA dataset is arguably the most commonly used benchmark for image-based pedestrian detection. It contains annotated image patches showing pedestrians scaled to a height of 96 pixels; all patches are padded by 12 pixels in four directions in order to provide contextual information. We therefore perform a statistical analysis with pedestrian images of size 60 × 120 pixels. On these data, we compute an average edge map based on gradient magnitudes extracted from each sample. The resulting average edge map is shown in Fig. 2 and clearly resembles a human body.

우리는 통계 정보를 기반으로 보행자 body shape 를 정의합니다. INRIA 데이터 세트는 이미지 기반 보행자 감지에 가장 일반적으로 사용되는 벤치마크 입니다. 보행자가 높이 96 픽셀로 조정된 주석 이미지 패치가 포함되어 있습니다. 문맥 정보를 제공하기 위해 모든 패치는 네 방향으로 12 픽셀 씩 패딩됩니다. 따라서 우리는 60 × 120 픽셀 크기의 보행자 이미지로 통계 분석을 수행합니다. 이 데이터에서 각 샘플에서 추출한 그래디언트 크기를 기반으로 평균 edge map 을 계산합니다. 결과적인 평균 edge map은 Fig. 2 그리고 분명히 인체와 닮았습니다.

Features derived from rectangular image regions typically allow for computational efficiency. We therefore decide to base our pedestrian detector on rectangular features and hence divide the edge map into square cells whose sizes may vary. Fig. 2 shows examples of cells of sizes 4 × 4 and 6 × 6 pixels. Given these grids of cells, the whole body is approximately divided into three parts: the head, the upper body, and the lower body. This is intended to increase robustness as these three parts generally appear in different colors or textures in real world images.

직사각형 이미지 영역에서 파생된 feature은 일반적으로 계산 효율을 허용합니다. 따라서 우리는 보행자 감지기를 직사각형 feature위에 놓기로 결정하고 edge map을 크기가 다를수 있는 정사각형 셀로 나눕니다. <Figure 2>는 크기가 4 × 4 및 6 × 6 인 셀들의 예를 보여준다. 이러한 grids of cells가 주어지면 body는 대략 head, upper body, lower body의 세 부분으로 나뉘어 집니다. 이 세 부분은 일반적으로 실제 이미지에서 서로 다른 색이나 텍스처로 나타나기 때문에 견고성을 높이기 위한 것입니다.

|  |
| --- |
|  |
| <Figure 2> Illustration of a statistical pedestrian shape model in terms of an average edge map as shown in the middle. In this example, cell sizes are chosen to be 4 × 4 and 6 × 6 pixels, respectively. Three bounding boxes approximately indicate the head, the upper body, and the lower body parts.  <그림 2> 중앙에 표시된 평균 에지 맵의 측면에서 통계 보행자 모양 모델의 그림. 이 예에서 셀 크기는 각각 4 × 4 및 6 × 6 픽셀로 선택됩니다. 세 bounding boxes는 대략 head, upper body, lower body 부분을 나타냅니다. |

**2.2. Generating templates**

**템플릿 제네레이팅**

We constrain our templates to be of rectangular form as these allow for convenient implementation and efficient computation. Statistical variations are coped with by considering different modalities as described in Sec. 3. First, however, we define a set of sizes

S = {(w, h) | w ≤ wm, h ≤ hm, w, h ∈ N+}, (1)

where w and h indicate the width and height (in terms of covered cells) of a rectangular template; wm and hm are used to constrain the overall size of templates since we focus on local image information.

우리는 템플릿을 직사각형 형태로 제한함으로써 편리한 템플릿 구현과 효율적인 계산이 가능합니다.

통계적 변이는 <섹션 3> 에서 기술된 것과 다른 양식을 고려함으로써 대응된다.

그러나 먼저, 일련의 크기를 정의하는데, 여기서 w와 h는 직사각형 틀의 너비와 높이 (덮여있는 셀의 관점에서)를 나타냅니다. wm과 hm은 로컬 이미지 정보에 초점을 맞추기 때문에 템플릿의 전체 크기를 제한하는 데 사용됩니다.

As shown in Fig. 2, images of pedestrians available in the INRIA data consist of four logical components: back-ground, head, upper body, and lower body. We assign each cell c(i, j) to a set of labels L(i, j) that indicate which components are found in the cell.

<Figure 2> 에서 볼 수 있듯이 INRIA 데이터에서 보행자 이미지는 배경, 머리, 상체, 하체의 네 가지 논리 구성 요소로 구성됩니다. 우리는 각 셀 c (i, j)를 셀에있는 구성 요소를 나타내는 레이블 L (i, j) 집합에 할당합니다.

Next, for each pair of sizes in S, we slide a corresponding rectangular window over the whole shape model to generate different templates at different positions and of different weights. At a certain position (x, y), the template to be created depends on how many different parts are contained in the rectangle. A binary template is generated if there are only two parts; ternary templates of different kinds are generated if there are three parts. Algorithm 1 provides details as to this procedure.

다음으로, S의 각 크기 쌍에 대해, 우리는 전체 모양 모델에 대해 대응하는 직사각형 창을 밀어서 다른 위치와 다른 무게의 다른 템플릿을 생성합니다. 특정 위치 (x, y)에서 생성 될 템플릿은 사각형에 몇 개의 다른 부분이 들어 있는지에 따라 달라집니다. 바이너리 템플릿은 두 부분으로 만 생성됩니다. 세 부분이 있으면 다른 종류의 3 진 템플릿이 생성됩니다. 알고리즘 1은 이 절차에 대한 세부 사항을 제공합니다.

The resulting full template pool is given as a set:

결과적인 전체 템플릿 풀은 아래의 <세트 T>로서 주어지며,

T = {(x, y, s, W) | x, y ∈ N, s ∈ S, W ∈ R2}, (2)

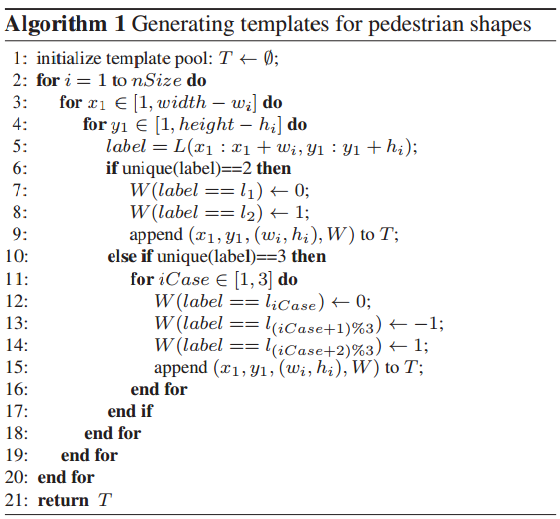
where x and y indicate the location of a template with respect to the human shape model and W is a weight matrix that is determined according to the matrix L of labels for all cells.

여기서 x 및 y는 인간 형상 모델에 대한 템플릿의 위치를 나타내고, W는 모든 셀에 대한 라벨의 매트릭스 L에 따라 결정되는 가중치 matrix이다.

**3. Multi-modal, multi-channel Haar features**

In the following, traditional Haar-like features will be referred to as binary modalities as they only carry two possible weights (+1 and −1) for different rectangles. However, this binary modality is ill suited to represent cusps or corner-like structures of the human silhouette. That is to say, that it hardly adapts to the description of the content of bounding boxes that contain three different logical components such as, say, head, upper body, and background. Yet, for efficient subsequent classification we are interested in computing the difference between parts with respect to two of them at a time. We therefore propose to consider ternary templates.

다음에서 전통적인 Haar-like 특징들은 다른 사각형들에 대해 두 개의 가능한 가중치 (+1과 -1)만을 가지고 있기 때문에 이진 양식 (binary modalities)이라고 할 것입니다. 그러나 이 바이너리 양식은 인간 실루엣의 첨단 또는 코너와 같은 구조를 나타내는 데 적합하지 않습니다. 말하자면 머리, 상체 및 배경과 같은 세 가지 논리 구성 요소를 포함하는 경계 상자의 내용에 대한 설명에 거의 적응하지 못합니다. 그러나 효율적인 후속 분류를 위해서는 한 번에 두 부품에 대한 부품 간의 차이를 계산하는 데 관심이 있습니다. 따라서 우리는 삼항 템플릿을 고려할 것을 제안합니다.



An example is given in Fig. 1 where ternary 2×2 templates capture the local geometry of the image region where head, shoulders, and background meet in joint corners.

To integrate color and gradient information, we build a multi-channel descriptor for each cell. We consider a total of 10 different channels as it is done in the detector [ChnFtrs]: 3 channels for LUV colors, 1 channel for gradient magnitude information, and 6 channels for histograms of oriented gradients.

<Figure 1>에서 ternary 2 × 2 템플릿은 머리, 어깨 및 배경이 모서리에서 만나는 이미지 영역의 local geometry를 캡처합니다.

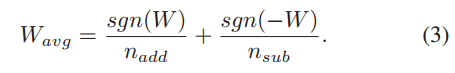
색상 및 그라디언트 정보를 통합하기 위해 각 셀에 대해 다중 채널 설명자를 작성합니다. 우리는 검출기 [ChnFtrs]에서 수행되는 것과 같이 총 10 개의 다른 채널을 고려합니다. LUV 색상의 경우 3 개, 그래디언트 크기 정보의 경우 1 개의 채널, 그래디언트의 히스토그램의 경우 6 개의 채널이 고려됩니다.

Assume we are given a template t = (x, y,(w, h), W).

t = (x, y, (w, h), W) 이라는 템플릿이 주어졌다 가정하자.

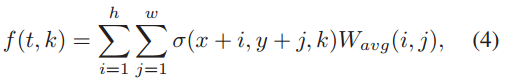
We first count how often the weights +1 and −1 appear and denote these counts as nadd and nsub. There are thus nadd additive cells and nsub subtractive cells and we normalize each cell’s weight by the total number of corresponding cells covered by a rectangle. This results in an average weight matrix:

먼저 가중치 +1과 -1이 나타나는 빈도를 계산하고 이 수를 nadd와 nsub로 나타냅니다. 따라서 nadd 첨가 셀과 nsub 빼기 셀이 있으며 각 셀의 가중치를 직사각형으로 덮여있는 해당 셀의 총 수로 표준화합니다. 평균 중량 행렬 결과 :



The feature value of any template t for any channel k, e.g. color or gradient information, can then be computed as a weighted sum:

임의의 채널 k에 대한 임의의 템플릿 t의 특징 값. 예를 들어, color 또는 그라디언트 정보는 아래와 같은 가중된 합으로 계산될 수 있으며 :



where, σ(i, j, k) denotes the sum of values in cell(i, j) along channel k.

여기서 σ(i, j, k)는 채널 k를 따르는 cell(i, j) 내의 값들의 합을 나타낸다.

|  |
| --- |
|  |
| Figure 3: Illustration of representative features.  (a) Cell weight map: different colors are used to indicate the accumulative weight of each cell after boosting.  (b) Most informative templates: these binary and ternary rectangle features obtained high accumulative weights after boosting.  그림 3 : 대표적인 feature 설명  (a) 셀 가중치 지도 : 증폭(boosting) 후 각 cell의 누적 가중치를 나타내기 위해 다른 색상이 사용됩니다.  (b) 가장 유익한 템플릿 : 이진 및 삼항 사각형 피쳐는(binary and ternary rectangle features) 증폭(boosting) 후 높은 누적 가중치를 얻었다. |

**4. Selecting features for pedestrian detection**

보행자 감지 feature 선택

Our detector employs the multi-modal and multi-channel Haar-like features proposed in <Section 3>. Note that these features are built on channel features as in [ChnFtrs], but interpret local differences between rectangular regions over multiple channels rather than over channel values themselves.

우리의 detector(검출기)는 <Section 3>에서 제안된 **multi-modal and multi-channel Haar-like features**를 사용합니다. 이러한 기능은 [ChnFtrs]에서와 같이 채널 feature를 기반으로 작성되지만 채널 value자체보다는 여러 채널에서 직사각형 영역 간의 로컬 차이점을 해석합니다.

We apply a fast version of AdaBoost [1] for learning since it offers a convenient and fast approach to select from a large number of candidate features. We apply 2048 decision trees of depth 2 to build our final strong classifier. Initial negative training samples are randomly generated and, afterwards, hard negative samples are searched for three rounds over all negative example images so as to collect 20,000 negative samples in total. This multi-round training strategy is pivotal as it leads to a better performance than a simple one round training procedure with the same number of negative samples. From our experiments, three rounds of retraining were observed to yield optimal performance; additional rounds did not show significant improvements.

AdaBoost [1]의 빠른 버전을 학습용으로 적용하여 다수의 후보 기능 중에서 편리하고 빠른 방법을 선택할 수 있습니다. 우리는 최종 깊이있는 분류자를 만들기 위해 깊이 2의 2048 개의 결정 트리를 적용합니다. 초기 네거티브 트레이닝 샘플을 무작위로 생성 한 후, 모든 네거티브 샘플 이미지에 대해 3 회 반복하여 하드 네거티브 샘플을 검색하여 총 20,000 개의 네거티브 샘플을 수집합니다. 이 **multi-round training strategy**는 같은 수의 negative 샘플을 가진 간단한 1 라운드 트레이닝 절차보다 더 나은 성능을 이끌어내는 데 중추적인 역할을합니다. 우리의 실험에서 최적의 성능을 내기 위해 3 회의 retraining(리트레이닝)이 관찰되었습니다. 추가 라운드에서는 큰 개선이 없었습니다.

In order to look into which features are more informative, we plot a weight image of the top 100 features as shown in <Figure 3a> To generate this figure, we add the weight of each feature to the cells it covers and use different colors to indicate the accumulative weight of each cell after boosting. As expected, the head-shoulder area of the human body shows to be more discriminative for pedestrian detection than other body parts.

어떤 특징이 더 유익한지 알아보기 위해 <Figure 3a>와 같이 상위 100 개 특징의 가중치 이미지를 그립니다.이 그림을 생성하려면 각 특징의 가중치를 포함하는 cell에 추가하고 다른 색상을 사용하십시오. boosting(증폭)후 각 셀의 누적 가중치를 나타냅니다. 예상대로, 인체의 머리 - 어깨 부분은 다른 신체 부위보다 보행자 감지에 더 차별적인 것으로 나타납니다.

The most discriminative binary and ternary templates determined by the boosting algorithm are then used for pedestrian detection in still images. To this end, we slide a window over the whole image and consider multiple scales. The spatial step size is set identical to the cell size for speed and the scale step is set to be 1.09 so that there are 8 scales in each octave. We use a simplified non-maximal suppression (NMS) procedure [8] to suppress nearby detections.

boosting(증폭) 알고리즘에 의해 결정되는 가장 차별적인 2진 및 3진 템플릿은 스틸 이미지의 보행자 디텍션에 사용됩니다. 이를 위해 우리는 전체 이미지 위에 창을 슬라이드하고 여러 개의 스케일을 고려합니다. 공간 step 사이즈는 속도에 대해 cell사이즈와 동일하게 설정되고 스케일 step은 1.09가되도록 설정되어 각 옥타브에 8 개의 눈금이 있습니다. 우리는 근사 검출을 억제하기 위해 **non-maximal suppression (NMS)** 절차 [8]를 사용합니다.

(\*\* **non-maximal suppression (NMS)** : image processing을 통해 얻은 edge를 **얇게** 만들어주는 것, 즉, 뭉개져 있는 edge를 더욱 **선명한 선**을 찾도록 진행해주는 것,  
참고사이트 : <https://m.blog.naver.com/jinsoo91zz/220511441402> )

**5. Experiments**

Experiments are conducted on two public benchmark datasets: the INRIA pedestrian dataset [5] and the Caltech pedestrian dataset [10]. The INRIA data is arguably the most popular dataset for people detection and comes along with pre-defined subsets for training and testing. The Caltech data is the largest and most challenging dataset for pedestrian detection and we consider subsets set00 - set05 for training and subsets set06 - set10 for testing.

Experiments는 INRIA 보행자 데이터 세트 [5]와 Caltech 보행자 데이터 세트 [10]의 두 가지 공개 벤치마크 데이터세트에서 수행됩니다. INRIA 데이터는 아마도 사람 감지를 위한 가장 인기있는 데이터 세트이며 교육 및 테스트를 위해 미리 정의된 하위 집합과 함께 제공됩니다. Caltech 데이터는 보행자 감지를 위한 가장 크고 도전적인 데이터세트이며 교육을 위해 set00 - set05 하위 집합을, 테스트를 위해 set06 - set10 하위 집합을 고려합니다.

**5.1. Implementation details**

**구현 세부 사항**

To optimize our detector, we analyze the influences of different parameter settings. Next, we present various experimental results on the INRIA dataset.

탐지기를 최적화하기 위해 다양한 매개 변수 설정의 영향을 분석합니다. 다음으로, INRIA 데이터 세트에 대한 다양한 실험 결과를 제시합니다.

Cell size: the pedestrian body shape can be covered by arrays of different cell sizes as shown in <Figure 2>. We present experimental results for cell sizes of 4 × 4, 6 × 6 and 8 × 8. From <Figure 4a>, we find that a cell size of 6×6 pixels produces the best results so we choose it as our default setting.

셀 크기 : 보행자의 몸 모양은 <Figure 2>와 같이 셀 크기가 다른 배열로 덮을 수 있습니다. 우리는 4 × 4, 6 × 6, 8 × 8 셀 크기에 대한 실험 결과를 제시한다. <Figure 4a>에서 6 × 6 픽셀의 셀 크기가 가장 좋은 결과를 얻음을 알 수있다.

Channels: we plot the performance of various channel combinations. As gradient histograms have been shown as the most informative channels in [8], we only try alternatives for color and gradient magnitude channels. From <Figure 4b> it appears that LUV color channels are more discriminative than HSV channels, both are commonly used in this area; using three gradient magnitude channels (one for each color channel) or two gradient components (along the x and y directions respectively) lead to slight decrease in performance rather than improvements.

채널 : 다양한 채널 조합의 퍼포먼스를 계획합니다. 그라데이션 히스토그램은 [8]에서 가장 유익한 채널로 표시되었으므로 색상 및 그라데이션 크기 채널에 대한 대안만 시도합니다. <Figure 4b>에서 LUV 컬러 채널은 HSV 채널보다 차별적인 것처럼 보입니다. 둘다 이 영역에서 일반적으로 사용됩니다. 3 개의 그라데이션 크기 채널 (각 색상 채널에 하나씩) 또는 두 개의 그라디언트 구성 요소 (x 및 y 방향으로 각각)를 사용하면 성능 향상보다는 성능이 약간 저하됩니다.

Image normalization: we analyze the influence of intensity normalization on our features as previous works on rectangular features typically employ various ways of normalization. [VJ] [24] used local normalization inside each detection window; [Roerei] [4] reported performance improvements by applying global normalization on the input images. However, according to the results in <Figure 4c>, our features obtain best results without normalization.

이미지 노멀라이제이션 : 직사각형 피쳐에 대한 이전 연구는 일반적으로 다양한 노멀라이제이션 방법을 사용하기 때문에 intensity normalization 이 우리 피쳐에 미치는 영향을 분석합니다. [VJ] [24] 각 탐지 창 내부에서 로컬 정규화를 사용했습니다. [Roerei] [4]는 입력 이미지에 전역 정규화를 적용하여 성능 향상을보고했습니다. 그러나 <Figure 4c>의 결과에 따르면 우리의 특징은 노멀라이제이션 없이 최상의 결과를 얻을 수 있다.

Smoothing: while pre-smoothing input images with binomial filters of radius 1 improves the performance by more than 3%, larger radii produce worse results; post-smoothing of channel features significantly decreases the performance and seems to inhibit characteristic local variations.

스무딩 : 반경 1의 이항 필터로 입력 이미지를 미리 평활화하는 것이 성능을 3 % 이상 향상 시키지만 반경이 클수록 반경이 더 커집니다. 채널 피처의 포스트 - 스무딩은 성능을 상당히 감소시키고 특징적인 국부적인 변동을 억제하는 것으로 보인다.

Number of weak classifiers: intuitively one would expect more weak classifiers to lead to better performance since decision boundaries become more accurate; on the other hand, too large number of weak classifiers may lead to overfitting of the training data. Accordingly, we find that detection performance starts to decrease slightly when the number of weak classifiers exceeds 2000.

weak classifiers의 수 : 직관적으로 결정 경계가 보다 정확해짐에 따라 더 약한 분류기가 더 나은 성능을 이끌 것으로 기대됩니다. 반면에, 너무 많은 수의 약한 분류기는 훈련 데이터의 과잉으로 이어질 수있다. 따라서 weak classifiers 수가 2000 개를 초과하면 검색 성능이 약간 떨어지기 시작합니다.

For the results reported next, we therefore consider the following settings of our detector: cell size of 6×6;channels of LUV+GM+GH; image smoothing with binomial filters of radius 1; no channel smoothing; no image normalization; 2000 weak classifiers.

따라서 위의 결과에서 우리는 탐지기의 다음 설정을 고려하게 되었다.

셀 크기 6 x 6 , LUV + GM + GH 채널 , 반경 1의 이항 필터로 이미지 평활화 , 채널 스무딩 없음 , 이미지 정상화 없음 , 2000 weak classifiers.

**5.2. Comparisons with state-of-the-art detectors**

**최첨단 탐지기와의 비교**

In this section, we compare our detector to other state-ofthe-art detectors whose results are publicly available. We use the same experimental protocol as in [10] and evaluate performances in terms of ROC curves. Measurements of average miss rates are used to summarize the overall performances of different detectors. The overall results are produced on the reasonable [10] subset of each test set which show pedestrians at a resolution of over 50 pixels in height and a visibility of at least 65%.

이 섹션에서는 우리의 detector를 공개적으로 이용가능한 다른 최신의 검출기와 비교합니다. 우리는 ROC 곡선의 관점에서 동일한 실험 프로토콜을 사용하고 성능을 평가합니다. 평균 누락률의 측정은 서로 다른 detector의 전반적인 성능을 요약하는 데 사용됩니다. 전체 결과는 높이가 50 픽셀 이상이고 가시성이 65 % 이상인 보행자를 보여주는 각 테스트세트의 합리적인 하위 집합에서 생성됩니다.

The results in <Figure 5a> show that our detector outperforms the baseline detector [ChnFtrs] by about 8% and reaches the state-of-the-art performance. The two detectors with better results than ours consider feature pools that are more than 20 times larger and are about 100 times slower in training.

<Figure 5a>의 결과는 우리의 detector 가 *베이스-라인 detector인 [ChnFtrs]* 보다 약 8 % 성능이 뛰어나고 최첨단 성능에 도달함을 보여줍니다. 우리보다 더 나은 결과를 보이는 두 개의 탐지기는 20 배 이상 더 커지고 훈련 속도가 약 100 배 느린 feature pool 을 고려합니다.

On the Caltech pedestrian dataset, our detector outperforms not only the baseline detector [ChnFtrs] by about 20% but also yields the overall best performance as shown in <Figure 5b> In particular, we note that it even outperforms detectors which consider additional motion information.

Caltech 보행자 데이터 세트에서, 우리의 detector 는 베이스 라인 검출기 [ChnFtrs]를 약 20 % 능가하지만 <Figure 5b>에서와 같이 전반적으로 최고의 성능을 보였다. 특히 우리는 추가적인 모션 정보를 고려한 detector 보다 성능이 뛰어나다 .

<Figure 6> shows evaluation results under different occlusion conditions for the Caltech pedestrian test data. As in [10], we use three occlusion levels: no occlusion (0% occluded), partial occlusion (1-35% occluded), and heavy occlusion (35%-80% occluded). The performance of all the detectors drops significantly as occlusion increases. Yet, our detector seems least affected by occlusion in the sense that it consistently ranks high for all occlusion levels. In fact, it achieves the best performance among all tested detectors for the cases of no and heavy occlusion and we conclude that the informed design of our features yields robustness against occlusions. Notably, our detector even outperforms those detectors that employ explicit occlusion handling strategies, for example [DBN-Isol] and [DBN-Mut], for all levels of occlusion.

<그림 6>은 Caltech 보행자 실험 데이터에 대한 다른 occlusion(오클루젼.. 잊어먹을만 하니 다시 나오네;;) 조건에서의 평가 결과를 보여준다. [10]에서와 같이 우리는 3 가지 occlusion 레벨을 사용합니다 : 오클루젼 없음 (0 % occluded), 부분 오클루젼 (1-35 % occluded) 및 헤비 오클루젼 (35 % -80 % occluded). 모든 detector의 퍼포먼스는 오클루젼이 증가함에 따라 크게 떨어집니다. 그러나 우리의 detector는 모든 오클루젼 수준에서 일관되게 높은 순위를 유지한다는 점에서 오클루젼에 가장 영향을 받지 않는 것으로 보입니다. 사실, 테스트된 모든 detector 중에서 오클루젼-없음 & 헤비-오클루젼 의 경우에 대해 최상의 성능을 얻었으며, 우리의 특징을 잘 알고 있으면 오클루젼 에 대한 견고성을 얻을 수 있다고 결론지었습니다. 특히, 우리의 detector는 모든 수준의 오클루젼 에 대해, 명시적인 오클루젼 처리 전략( explicit occlusion handling strategies) 을 사용하는 detector (예 : [DBN-Isol] 및 [DBN-Mut]) 보다 우수 합니다

**5.3. Feature size and runtimes**

**피쳐 사이즈 & 런타임**

We present our feature size with the optimal settings concluded from <Section 5.1>.

우리는 <Section 5.1>. 에서 결론 지은 최적의 설정으로 피처-사이즈 를 제시합니다.

Given 6 × 6 cells and templates size ranging from 1 × 2 to 4 × 3 cells, we obtain 266 templates at different positions.

6 × 6 cells 와 1 × 2에서 4 × 3 cell 범위의 템플릿 크기를 감안할 때, 우리는 다른 위치에서 266 개의 템플릿을 얻습니다.

Shifting templates along 4 directions with a step of one cell yields a template pool of 1276 (some shifts are not possible at image borders); considering 10 channels, the final feature size is 12,760.

한 cell의 단계로 4 방향을 따라 템플릿을 이동하면 1276의 템플릿 풀이 생성됩니다 (이미지 테두리에서는 일부 교대가 불가능합니다).

10 개의 채널을 고려할 때 최종 피쳐 크기는 12,760입니다.

Our detector is implemented in Matlab, on an Intel Core i7 CPU (3.5GHz). On the Caltech dataset, it takes 1 hour for training with 4 rounds and 1.6 seconds ( [ChnFtrs] 2s) for testing a 640 × 480 image using the optimal parameters as illustrated in Sec. 5.1. In addition to channel computation, our feature computation includes local sums and differencing, both of which can be parallelized for further speed-up. Our detector is expected to reach real-time efficiency running on a powerful machine and with GPU computation enabled.

우리의 detector는 Intel Core i7 CPU (3.5GHz)의 Matlab에서 구현됩니다. Caltech 데이터 세트에서 Sec에 설명된대로 최적의 매개 변수를 사용하여 640 × 480 이미지를 테스트하는 데 4 회 및 1.6 초 ([ChnFtrs] 2s)로 교육하는 데 1 시간이 소요됩니다. 5.1. 채널 계산 외에도 우리의 피쳐 계산에는 지역 합계와 차이 계산이 포함되며, 두 가지 모두 병렬 처리하여 속도를 향상시킬 수 있습니다. 우리의 detector 는 강력한 기계에서 실행되고 GPU 계산이 가능한 실시간 효율에 도달할 것으로 예상됩니다.

|  |
| --- |
|  |
| Figure 4: Evaluation of different parameters on the INRIA pedestrian dataset.  (a) Cell sizes of the pedestrian shape model.  (b) Channel combinations with color channels + gradient magnitude channels (GM) + gradient histogram channels (GH).  (c) Image normalization methods. Local intensity normalization is done inside each detection window; global normalization is done for the whole input image.  (d) Pre-smoothing of colors with binomial filters of different radii.  (e) Post-smoothing of channels with binomial filters of different radii.  (f) Number of weak classifiers.  그림 4 : INRIA 보행자 데이터 세트의 다양한 매개 변수 평가.  (a) 보행자 형상 모델의 셀 크기.  (b) 색상 채널 + 경사도 크기 채널 (GM) + 경사 막대 그래프 채널 (GH)의 채널 조합.  (c) 이미지 정규화 방법. Local intensity normalization은 각 Detection window 내에서 수행됩니다. global normalization(전역 정규화)는 전체 입력 이미지에 대해 수행됩니다.  (d) 다른 반경의 이항 필터로 색을 Pre-smoothing.  (e) 반경이 다른 이항 필터로 채널의 Post-smoothing.  (f) weak classifiers의 수. |
|  |
| Table 1: Performance comparisons for state-of-the-art pedestrian detectors.  Each row in this table summarizes information as to features and classifiers used in a particular approach, and displays the corresponding average performance in terms of miss rates. The approach proposed in this paper yields state-of-the-art performance on the INRIA dataset and consistently better results than previously reported on the Caltech dataset.  Table 1: 최첨단 보행자 detector의 성능 비교.  이 표의 각 행은 특정 접근 방식에서 사용되는 feature 및 classifier 에 대한 정보를 요약하고 해당 평균 성능을 실패율로 표시합니다. 이 백서에서 제안 된 접근법은 INRIA 데이터 세트에 최첨단 성능을 제공하고 Caltech 데이터세트에서 이전에 보고 된 것보다 지속적으로 더 나은 결과를 산출합니다. |

|  |
| --- |
|  |
|  |

**6. Conclusion**

**결론**

We considered the problem of efficient yet robust pedestrian detection from image data. The particular approach we presented in this paper was motivated by the observation that a current trend in work on pedestrian detection consists in analyzing feature vectors of ever increasing dimensions which necessitate the use of powerful hardware in order to guarantee real time capability.

우리는 이미지 데이터로부터 효율적이지만 강력한 보행자 감지 문제를 고려했다. 이 논문에서 제시한 특별한 접근 방식은 실시간 능력을 보장하기 위해 강력한 하드웨어의 사용을 필요로 하는 점점 더 커지고 있는 차원의 특징 벡터 분석에 보행자 탐지 작업의 현재 추세가 있다는 관찰에 의해 동기 부여되었습니다.

Also, because of the peculiar geometry of high dimensional spaces (concentration of measure and neighborliness) it is not necessarily guaranteed that additional efforts spent on computing high dimensions pay off in terms of recognition accuracies. We therefore explored more compact features could yield state-of-the-art performance in pedestrian detection if they were designed based on prior information as to the appearance of the up-right human body.

또한, 높은 차원 공간(측정 및 인접성의 집중)의 독특한 기하학으로 인해, 높은 차원을 계산하는데 소비되는 추가 노력이 인식 정확도 측면에서 돋보이는 것은 반드시 보장되지는 않는다. 따라서 우리는 보행자 탐지가 최신의 인체 모습과 관련하여 사전 정보를 기반으로 설계된 경우보다 콤팩트한 기능이 보행자 탐지에서 최첨단 성능을 낼 수 있다고 생각했습니다.

Given a large dataset of pedestrian images, we computed a statistical shape model which proved to consist of four clearly recognizable logical components. We covered this shape model with grids of cells and slid rectangular windows over these cell arrays to produce a set of location specific weighted binary or ternary Haar-like templates that incorporate information as to which of the four components of the shape are covered by a rectangle.

보행자 이미지의 큰 데이터 세트가 주어지면, 우리는 4 개의 명확하게 인식할 수있는 논리적 구성 요소로 구성된 것으로 입증된 통계적 형상 모델을 계산했습니다. 우리는 cell의 grid로 이 shape 모델을 다루었고 이 셀 배열 위에 직사각형 윈도우를 밀어 넣어 모양의 4 개 구성 요소 중 어느 사각형이 직사각형으로 덮여 있는지에 대한 정보를 통합한 위치 특정 가중 바이너리 또는 3 개 Haar 형 템플릿을 생성했습니다

The weighting scheme provided us with a simple mechanism of generating multi-modal & multi-channel Haar-like features and we applied boosting to determine the most informative ones. As our approach does not require computing any possible configuration of rectangles within a sliding window nor is based on random sampling of rectangle features, it marks a middle ground among recently published similar approaches. Moreover, our detector is inherently simple to implement, easy to train, and fast during runtime.

가중치 체계는 다중 모드 및 다중 채널 Haar와 유사한 기능을 생성하는 간단한 메커니즘을 제공하고 가장 유익한 기능을 결정하기 위해 부스팅을 적용했습니다. 우리의 접근법은 슬라이딩 윈도우 내에서 직사각형의 가능한 구성을 계산할 필요가 없으며 직사각형 피쳐의 무작위 샘플링을 기반으로하므로 최근에 발표 된 유사한 접근 방식 중 중간에 표시됩니다. 또한, 우리의 탐지기는 본질적으로 구현이 간단하고, 훈련이 쉽고, 실행 중에 빠르다.

In extensive experiments with standard benchmark datasets, we found our detector to achieve state-of-the-art performance on the INRIA pedestrian dataset and, for the Caltech pedestrian dataset, we found it to outperform all other recent approaches considered in our tests. In addition, our model-based rectangular features proved to be highly robust under occlusion and even outperformed methods that contain explicit mechanisms for occlusion handling.

표준 벤치 마크 데이터 세트를 이용한 광범위한 실험에서 탐지기가 INRIA 보행자 데이터 세트에 최첨단 성능을 제공하고 Caltech 보행자 데이터 세트에 대해 우리가 테스트에서 고려한 다른 모든 최근 접근법을 능가하는 것으로 나타났습니다. 또한 모델 기반 직사각형 피처는 오클루젼 하에서 매우 견고하고 오클루전 처리를 위한 명시적인 메커니즘을 포함하는 성능이 우수한 방법임이 입증되었습니다.

Given these results, it appears promising to further explore model driven design of efficient rectangular features. Immediate extensions of the approach presented in this paper could be to incorporate additional channels such as motion information. More challenging extensions consist in adapting our scheme to scenarios where the objects to be detected show higher intra-class variations.

이러한 결과를 감안할 때 효율적인 직사각형 피쳐의 모델 기반 디자인을 더 연구하는 것이 유망한 것으로 보입니다. 이 백서에 제시된 접근법을 즉각적으로 확장하면 동작 정보와 같은 추가 채널을 통합 할 수 있습니다. 보다 도전적인 확장은 탐지될 객체가 더 높은 클래스 내 변화를 보이는 시나리오에 우리의 계획을 적용하는 것으로 구성됩니다.

|  |
| --- |
|  |
| Figure 6: Evaluation results under different occlusion conditions on the Caltech pedestrian test dataset. Caltech 보행자 테스트 데이터-세트의 다른 오클루젼 조건 하에서의 평가 결과. |

**References**

[1] R. Appel, T. Fuchs, P. Dollar, and P. Perona. Quickly boost- ´ ing decision trees-pruning underachieving features early. In ICML, 2013. 4

[2] A. Bar-Hillel, D. Levi, E. Krupka, and C. Goldberg. Partbased feature synthesis for human detection. In ECCV, 2010. 6

[3] R. Benenson, M. Mathias, R. Timofte, and L. V. Gool. Pedestrian detection at 100 frames per second. In CVPR, 2012. 2, 6

[4] R. Benenson, M. Mathias, T. Tuytelaars, and L. V. Gool. Seeking the strongest rigid detector. In CVPR, 2013. 2, 5, 6 [5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005. 2, 5, 6

[6] P. Dollar, R. Appel, and W. Kienzle. Crosstalk cascades for ´ frame-rate pedestrian detection. In CVPR, 2012. 2, 6

[7] P. Dollar and P. Perona. The fastest pedestrian detector in the ´ west. In BMVC, 2010. 2

[8] P. Dollar, Z. Tu, P. Perona, and S. Belongie. Integral channel ´ features. In BMVC, 2009. 2, 5, 6

[9] P. Dollar, Z. Tu, H. Tao, and S. Belongie. Feature mining for image classification. In CVPR, 2007. 2

[10] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian ´ detection: an evaluation of the state of the art. IEEE Trans. PAMI, 34(4):743–761, 2011. 1, 5

[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained partbased models. IEEE Trans. PAMI, 32(9):1627–1645, 2010. 2, 6

[12] P. F. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In CVPR, 2008. 2, 6

[13] J. Gall, A. Yao, N. Razavi, L. V. Gool, and V. Lempitsky. Hough forests for object detection, tracking, and action recognition. IEEE Trans. PAMI, 33(11):2188–2202, 2011. 2

[14] D. Levi, S. Silberstein, and A. Bar-Hillel. Fast multiple-part based object detection using kd-ferns. In CVPR, 2013. 6

[15] J. J. Lim, C. L. Zitnick, and P. Dollar. Sketch tokens: a ´ learned mid-level representation for contour and object detection. In CVPR, 2013. 2, 6

[16] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In CVPR, 2008. 6

[17] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In CVPR, 1997. 2

[18] W. Ouyang and X. Wang. A discriminative deep model for pedestrian detection with occlusion handling. In CVPR, 2012. 6

[19] W. Ouyang, X. Zeng, and X. Wang. Modeling mutual visibility relationship with a deep model in pedestrian detection. In CVPR, 2013. 6

[20] C. Papageorgiou and T. Poggio. A trainable system for object detection. IJCV, 38(1):15–33, 2000. 2

[21] D. Park, D. Ramanan, and C. Fowlkes. Multiresolution models for object detection. In ECCV, 2010. 6

[22] D. Park, C. L. Zitnick, D. Ramanan, and P. Dollar. Exploring ´ weak stabilization for motion feature extraction. In CVPR, 2013. 6

[23] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. In CVPR, 2007. 6

[24] P. Viola and M. J. Jones. Robust real-time face detection. IJCV, 57(2):137–154, 2004. 2, 5, 6

[25] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. In CVPR, 2010. 2, 6

[26] X. Wang and T. X. Han. An HOG-LBP human detector with partial occlusion handling. In ICCV, 2009. 2, 6

[27] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In CVPR, 2009. 1

[28] J. Yan, X. Zhang, Z. Lei, S. Liao, and S. Z. Li. Robust multiresolution pedestrian detection in traffic scenes. In CVPR, 2013. 6