Latent Subcategory Models for Pedestrian Detection with Partial Occlusion Handling

Samuele Martelli[1], Marco San Biagio[1] and Vittorio Murino[1,2]

[1] Pattern Analysis & Computer Vision (PAVIS) - Istituto Italiano di Tecnologia, Genova, Italy
[2] Dipartimento di Informatica - Università degli Studi di Verona, Verona, Italy
{Name.Surname}@iit.it

Abstract

Pedestrian detection is one of the most important tasks in Computer Vision, especially in automotive and security applications. One of the most common problems in real scenarios is related to the detection of occluded pedestrians. In this paper, we propose a novel multi-cue pedestrian detection approach able to deal with non homogeneous object samples by learning latent subcategory models trained on both visual and depth-based features. We also propose a novel self-similarity based feature, namely SST_D, to encode the homogeneity in appearance of pedestrians characterized by similar occlusion patterns. Experiments are performed on the Daimler Pedestrian Detection Benchmark Dataset showing the robustness of our approach in actual scenarios.

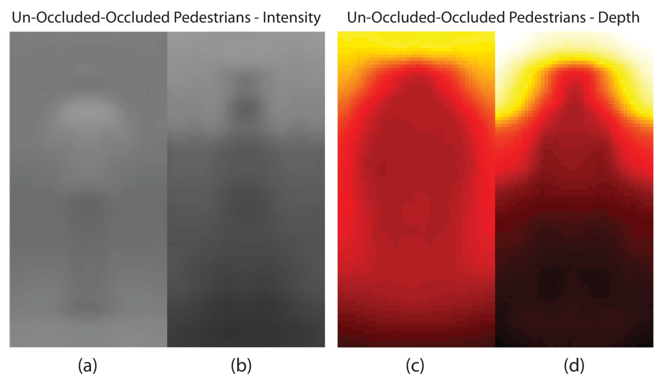

Figure 1. Average over a sub-set of images from [6] in the case of un-occluded and occluded pedestrians for intensity and depth maps. The human silhouette is clearly visible in (a-c), while it is occluded in (b-d).

1. Introduction

Pedestrian detection is an important and complex task in Computer Vision, representing one of the most basic operations in many significant applications such as automotive, video-surveillance, robotics and content-based image/video retrieval, to cite a few [5, 6]. The articulated structure and variable appearance of the human body, combined with scarce illumination, poor image resolution, pose variations and different point of views, contribute to increase the complexity of the problem. Furthermore, in case of a moving camera in a dynamic environment, changing backgrounds and partial occlusions may cause additional problems.

In the literature, the standard strategy consists in the use of a sliding windows approach over the whole image and the use of a classifier to evaluate each window. The classifier may be trained with different features, *e.g.* Haar-like features [24], Histograms of Oriented Gradients (HOG) [2], binary descriptors [26] or relation-based descriptors [19]. Popular classifiers are Support Vector Machines (SVMs) or boosting-based algorithms assembled in rejection cascades (*e.g.* AdaBoost [24], LogitBoost [13]). For a complete survey on pedestrian classification see [1, 5].

Among them, one of the most successful methods is the one presented in [2], based on an SVM classifier trained with HOG features. However, the main drawback of this approach is that it is based on a monolithic object-model representation, thus suffering in those scenarios where the objects of interest are characterized by articulated poses and they are observed by a perspective point of view.

To partially solve this issue, part-based approaches, which represent a pedestrian as an ensemble of parts, have been recently proposed such as the popular Deformable Part Model (DPM) [12]. It has also drawn the attention from the entire Computer Vision community becoming part of classification, segmentation, tracking and action recognition systems. In general, it performs better than HOG detector, because it has a richer object model defined by a global "root" filter enriched by several deformable part templates, trained on a latent discriminative learning framework.

The original DPM has also been extended to handle multiple views (subcategories) [11, 4], and to encode asymmetric patterns [9]. Detection accuracy has been further increased using grammar models [14]. Training and testing times have been reduced by using latent Linear Discriminant Analysis (LDA) [15] and partial hypothesis pruning [10], respectively. Overall, the latest version has reached state-of-the-art performance in challenging object detection datasets outperforming about three times the average precision of the original HOG detector. However, as pointed out in [5], in real-world scenarios such as automotive and video surveillance, images are often characterized by low-resolution so that the benefits of parts in the DPM become negligible. In general, pedestrian detection still remain a challenging task. Even latest state-of-the-art detectors do not yet guarantee the robustness and classification performance required for their deployment into real-world systems.

In this context, an emerging trend is the exploitation of multiple imaging modalities, *e.g.* combining optical cameras with other sensors such as thermal, infrared or stereo cameras. For example, depth-based computer vision systems have revealed their effectiveness in different tasks like pose estimation [20] and tracking [21]. Depth-based methods have recently proved to be significantly more efficient also in the pedestrian detection task [6, 25, 7, 17, 16]. In particular, a multi-cue pedestrian detection system with partial occlusion handling has been presented in [6]. The method has been mainly developed to reach low false positive rates required in automotive applications, combining visual, depth and motion features. The main drawback of this approach is that occluded pedestrian patterns are not taken into account while training the object classifier. Partial occlusion handling is limited to the test phase, by using an heuristic calculated on the depth and motion maps.

On the contrary, it has been recently demonstrated that partial occlusion modelling at the classifier level is very effective, even using only optical images [23]. Indeed, as pointed out in [5] and illustrated in Fig.1, pedestrians are often partially occluded as they move in proximity of other objects, especially in outdoor scenes. As a consequence, occlusions can significantly affect the overall performance of the detector if they are not properly considered [5].

In this paper, we propose a novel multi-cue pedestrian detection approach able to deal with significant appearance variations (*i.e.* occlusions) by learning a latent subcategory model trained on both visual and depth-based features. Each subcategory classifier encodes a different characteristic of the object class, such as a particular occlusion pattern. We also propose a novel self-similarity based feature, namely SST_D, encoding the human depth silhouette to boost the performance even further. In the experiments, first we empirically show the benefits of modelling occlusions in the training phase by incrementally feeding additional partially occluded examples into a standard linear SVM classifier. Second, by modelling occlusion patterns with subcategory classifiers and encoding the human silhouette with the novel self-similarity feature, we set state-of-the-art performance on the challenging Daimler Pedestrian Detection Benchmark Dataset [6].

The rest of the paper is organized as follows. The proposed method is described in Sec. 2, introducing the latent subcategory models and the self-similarity based feature. Experiments in Sec. 3 report the comparative performances of the proposed method and, finally, conclusive remarks and future work is discussed in Sec. 4.

2. The proposed Method

Training a single global classifier for a highly variable object class such as pedestrians, usually leads in a low complexity model which underfits the data. On the contrary, learning latent subcategory models provide significant advantages because they can cope with finer object variations in a principled way. Inspired by the recent work of [4], we modified the original framework of the DPM [11] in order to represents people by a compositional model encoding the appearance of highly variable instances. As a result of the training process, each subcategory model is specialized in a particular object category, in particular pedestrians occluded by the same occlusion pattern (*e.g.* pedestrians occluded in the lower part of the image). In Sec.2.1, we briefly review the key aspects of using subcategory models and in Sec.2.2 we explain in detail how this approach has been extended to handle partial occlusions in a multi-cue depth-based framework.

2.1. Latent Subcategory Models

Given a set $X = (x_1, x_2, \dots x_n)$ of n annotated instances of an object (*e.g.* pedestrians and background images) and their corresponding labels $y_i \in \{-1, 1\}$, in the training phase a set of K binary classifiers (subcategory models) is learned, wherein each individual classifier is trained on different subsets of the training data. Each subcategory model is defined as a global classifier, called "root", and several part classifiers whose displacements with respect to the root is encoded as a latent variable z.

If using standard linear SVMs for both the root and part classifiers, the pedestrian detection problem is reduced to a binary classification task (*e.g.* pedestrian or not) formulated as the following Latent-SVM (LSVM) optimization problem:

$$\underset{w}{\operatorname{argmin}} \frac{1}{2} \sum_{k=1}^{K} \|w_k\|^2 + C \sum_{i=1}^{n} \epsilon_i, \tag{1}$$

$$y_i \cdot s_i^{z_i} \geq 1 - \epsilon_i, \epsilon_i \geq 0, \tag{2}$$

$$z_i = arg \max_k s_i^k, \tag{3}$$

$$s_i^k = w_k \cdot \varphi_k(x_i) + b_k. \tag{4}$$

The learned model is defined by $W = (w_0, w_1, \ldots, w_{K-1})$, where w_k denotes the separating hyperplane for the $k-$th subcategory learned on the feature space defined by the feature mapping function φ (usually HOG). The learning process depends on the parameter C that controls the relative weight of the hinge-loss term.

Since the minimization function is semi-convex, the model parameters w_k and the latent variables z are learned using an iterative approach [11]. It is well known that a key step for the success of this latent subcategory approach is to generate a good initialization. It has been shown in [4] that the original semantic clustering scheme used for subcategory assignment [11], based on instance aspect-ratios (human annotated bounding-boxes), is prone to local minima especially when a large number of K subcategories is chosen. As a consequence, the original implementation of DPM [9] was limited to a compositional model of only 3 subcategories (components).

Instead, we experimentally noticed that by using a different partitioning scheme such as the data driven clustering proposed in [4], the number of subcategories K can be extended increasing the detection performance. Furthermore, it has been noticed that the contribution of the deformable parts become negligible by increasing the number of components [4]. As a result, deformable parts can be removed and the set of latent variables z can be consequently restricted to a single variable which specifies the subcategory assignment. The resulting mixture model is called Latent Subcategory Models (LSM) and represents a special case of Latent SVMs. It is faster in training, testing and less prone to local minima [4].

2.2. Encoding Occlusions with Self-Similarities

In order to boost the performance ever further, the proposed classification model can be trained with a better object representation φ, alternative to the standard HOG feature. In particular, we propose a novel self-similarity paradigm specifically designed to encode occlusion patterns by comparing the statistical distribution in terms of depth between different human body parts.

Our assumption is that in a typical automotive or video-surveillance setting, parts of un-occluded pedestrians are approximately at the same distance from the camera (assuming pedestrians standing upright on the ground) as illustrated in Fig.1(c). Different body parts are characterized by heterogeneous appearance but the depth distribution is almost the same. Statistical evidence shows that the head, shoulders and torso of un-occluded pedestrians are more correlated in terms of depth than legs or arms (Fig.1(c)). Also some background regions around un-occluded pedestrians are correlated, corresponding for example to the ground floor. Instead, in the presence of partial occlusions, some parts do not respect this assumption any more. However, the structure of a pedestrian is still preserved in the other areas, usually the upper body parts (see Fig.1(d)). Our contribution is to encode this information by using a self-similarity based structural descriptor, fed into the latent subcategory framework.

Structural descriptors, such as those in [3], have been usually applied to object detection using optical images. Similarity-based representations are mostly suited for encoding structural information. Similarities explain an object in relation to how much it is similar to other reference entities. In case of self-similarity, each part of an object can be defined in relation to the others, by considering proper distance (or dissimilarity) measures. The structure of the object as a whole can be extracted as mutual relations between parts.

One of the few approaches that successfully applied the self-similarity paradigm to depth data is the work presented in [17], where the Self-Similarity Tensor (SST) was proposed. The original SST feature is built through pairwise comparisons of HOG features [2], computed over the depth map. However, contrast-based features such as HOG are not well suited for the noisy and relatively smooth depth maps. Instead, we apply this paradigm directly comparing statistical distributions of depth calculated on regions (blocks) inside the detection window, and we call it SST_D.

In short, SST_D features are calculated in three steps. Given the bounding box B with the same size of the detection window where the object of interest is enclosed, first, we apply a quantization procedure on the depth map D. We a-priori define a range corresponding to the search area in the 3D camera set-up configuration by setting a minimum and a maximum depth value, respectively equal to d_{min} and d_{max} (*e.g.* 10-50 meters).

Then, given a depth map D, we define a regular grid of W square patches of size S pixels. The pixel values of the patches are collected into the vector

$$\mathbf{\Lambda}(\mathbf{D}) = [\, P_1(D),\ P_2(D), \ldots, P_W(D)], \tag{5}$$

where P is the patch extractor operator, producing an S-dimensional feature vector.

For each patch, P_i, we compute a local histogram, $\mathbf{H}_i$, resulting in

$$\mathbf{H}(\mathbf{D}) = [\, \mathbf{H}_1,\ \mathbf{H}_2, \ldots, \mathbf{H}_W]. \tag{6}$$

The histogram $\mathbf{H}_i$ is calculated using a fixed bin resolution d_Δ, resulting in a patch histogram of size h. In practice, the resulting $\mathbf{H}(\mathbf{D})$ vector of size $W \times h$ collects the local statistics, each one approximated by the local histogram $\mathbf{H}_i$.

In the second step, local information encoded in the patches is combined by computing similarity measures between histograms. In this way, highly invariant structural knowledge can be distilled. We compare every possible pair of region descriptors (histograms), and each comparison is encoded as the distance between the two histograms. The similarity/dissimilarity between two patches is computed as:

$$d_{i,j} = \delta(\mathbf{H}_i, \mathbf{H}_j) \;\; i,j = 1, \dots W. \tag{7}$$

where δ is the distance function. In this work we set δ as the Histogram Intersection Kernel (HIK) [22], but any distance function (*e.g.* Euclidean distance) can be adopted.

In the final step, applying (7) to all the possible pairs of patches included in (5), we extract the SST_D descriptor, defined as the symmetric matrix:

$$SST_D = \begin{bmatrix} d_{1,1} & \cdots & d_{1,W} \\ \vdots & \ddots & \vdots \\ d_{W,1} & \cdots & d_{W,W} \end{bmatrix}, \tag{8}$$

In case δ is non directional (*e.g.* HIK, L1), it is possible to vectorize the upper triangular sub-matrix without losing information, obtaining $SST_D \in \mathbb{R}^{W(W+1)/2}$

$$SST_D = [d_{1,1} \; d_{1,2} \; \dots d_{1,w} \;\; d_{2,2} \; \dots \; d_{W,W}]. \tag{9}$$

It is worth noting that SST_D is fast to compute because does not require heavy gradient and Gaussian re-weighted computations as [17], and it takes advantage of the integral histogram representation [18]. The SST_D tensor is a sort of *structural characterization* of the object of interest (see Fig.3). This feature is here employed to encode the depth silhouette characterizing human poses and it has shown to be very effective to encode partial occlusions as well, as described in the next section.

3. Experiments

In this section we present the results obtained with the proposed method tackling the challenging problem of pedestrian detection in automotive and video-surveillance scenarios.

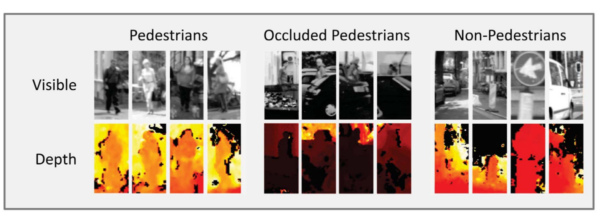

Figure 2. Positive (occluded and un-occluded) and negative examples extracted from the Daimler Multi-Cue dataset.

We carried out our experiments on a challenging dataset, named *Daimler Pedestrian Detection Benchmark Dataset* [6]. Some examples extracted from this dataset are shown in Fig. 2. The Daimler Pedestrian Detection Benchmark Dataset contains 77720 un-occluded positive, 11160 partially occluded and 48700 negative samples. This dataset is captured from a moving vehicle in complex urban traffic. Training and testing samples have a resolution of 48×96 pixels with a 12-pixel border around the pedestrians. The training set is split into 52122 positive and 32465 negative samples, while the test set consists of 25608 positive and 16235 negative samples. Each image is composed of three image modalities: standard visible grey scale image, depth, and motion flow. In our experiments, we take into account only the gray scale and depth images.

This section is subdivided into three parts: in sub-section 3.1 the object model used for the experiments is presented. In Sub-section 3.2, we show an analysis of the results obtained when an increasing number of occluded pedestrians are fed into the training phase of a standard HOG detector with a linear SVM. Furthermore, we present the advantages of our DPM implementation in a multi-cue setting using the SST_D. Finally, Sub-section 3.3 presents a detailed analysis of the proposed method with different configuration parameters.

3.1. Object Model

Considering an image I, we extract the HOG descriptor [2] on a predefined number of patches N, of size $width \times height$, with an overlap of $width/2$ and $height/2$ in the row/column dimensions, following the same procedure of [6]. Instead, the SST_D descriptor is extracted following the procedure described in Sec.2. Fig.3 shows the object model used in all the following experiments.

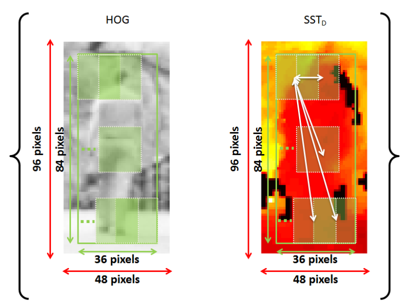

Figure 3. Final Feature Vector. Representation of the computation of $HOG_I + SST_D$.

3.2. Modeling Partial-Occlusions

To evaluate the performances of our system, we used the ROC curve which measures the detection performance by true positive rate over false positive per window. As first experiment, we carried out an evaluation of a standard HOG descriptor [2] computed on both intensity and depth images (HOG_{I+D}) coupled with a linear SVM classifier fed with the un-occluded training set and an increasing number of

Components K	only Un-Occluded		only Occluded		Un-Occluded+Occluded	
	SVM	LSVM	SVM	LSVM	SVM	LSVM
1	96.63%	-	92.31%	-	94.47%	-
3	96.66%	96.83%	94.33%	95.61%	95.50%	96.22%
5	96.82%	96.87%	94.45%	94.89%	95.63%	95.88%
7	96.61%	96.59%	94.80%	**97.24%**	95.71%	**96.91%**
10	96.49%	**96.91%**	95.20%	94.57%	95.85%	95.74%
15	96.56%	96.80%	93.75%	95.86%	95.15%	96.33%
30	96.48%	96.62%	96.28%	95.05%	96.38%	95.84%

Table 1. Evaluation of HOG_{I+D} changing the number of K subcategories with SVM and LSVM classifiers.

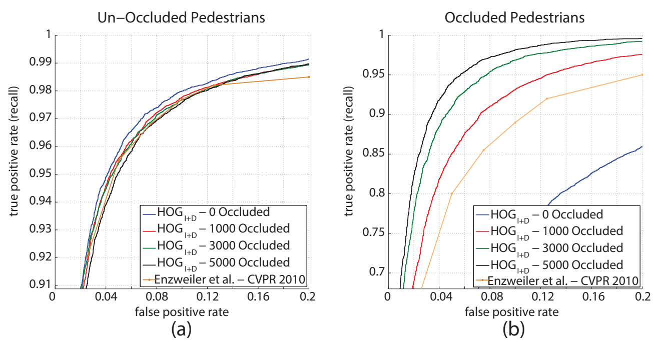

Figure 4. Performances evaluation obtained on the Daimler Pedestrian Detection Benchmark Dataset [6] increasing number of occluded examples in the training phase in the case of (a) unoccluded (b) occluded pedestrians.

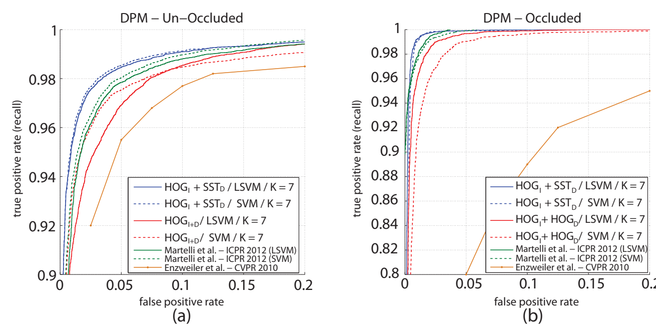

Figure 5. Performances evaluation obtained on the Daimler Pedestrian Detection Benchmark Dataset [6] using the DPM approach. Classification performances on (a) un-occluded and (b) occluded pedestrians.

occluded examples (0, 1000, 3000 and 5000). As baseline, we used the state-of-the-art multi-cue approach proposed in [6]. Performances on Un-Occluded pedestrians with HOG_{I+D} are showed in Fig.4(a). As expected, increasing the number of occluded pedestrians in the training set the performance slightly decreases while remaining quite stable and comparable with [6]. On the contrary, Fig. 4(b) shows that performance on the Occluded Pedestrian dataset significantly increases as more occluded examples are fed into the classifier. It is also interesting to notice that our method clearly outperforms the state-of-the-art result of [6] when only a relatively small number of 1000 occluded samples (red line in Fig. 4(b)) are included into the training set, composed of 52122 un-occluded samples.

In the second experiment, we tested the performances of our approach using a linear SVM and latent subcategory models (LSVM) with different combinations of features, extracted from intensity (HOG_I [2]) and depth maps (HOG_D, SST [17], SST_D). In this experiment, the number of subcategories has been fixed equal to 7 and 3000 occluded examples have been included in the training dataset. Fig. 5 shows the results obtained in the case of (a) unoccluded and (b) occluded pedestrians. As one can notice, using SST_D in combination with HOG_I we achieved the best performance in both datasets, overcoming the previous state-of-the-art results in [6]. This experiment also demonstrates that SST_D is more effective than HOG_D and SST to encode the humans' depth silhouette and occlusion patterns. As expected, the benefits of using latent subcategory models with respect to a single SVM classifier are more evident with occluded pedestrians (Fig. 5(b)), because of the high variability induced by different occlusion patterns. Furthermore, the benefits of a better classification framework (LSVM) are more significant when less discriminative features such as HOG_D are adopted (see LSVM vs. SVM, red lines in Fig. 5(a-b)).

The outcome of these experiments is then two-fold. First, we find out that (i) by including only a relative small portion of partially occluded samples into the training dataset, and (ii) handling occlusions by learning latent subcategory models, considerable improvement in performance is obtained. These observations suggest that complex heuristic-based methods to handle occlusions can be overcome by conceptually simpler methods using robust machine learning frameworks.

3.3. Performance Analysis

To provide more comprehensive analysis of the proposed method, we evaluated its performance with respect to the number of subcategories K. Detailed analysis comparing the performances of our implementation using HOG_{I+D} are reported in Table 1. The table shows the AP (Average Precision) values computed with the PASCAL VOC criteria [8] for both SVM and LSVM.

The first column of Table 1 represents the number of centroids, or subcategories, considered in the k-means algorithm. The other columns correspond to the classification results obtained with SVM and LSVM in the case of un-occluded, occluded and un-occluded+occluded pedestrians, respectively (in the latter case the final result is given averaging the other two corresponding results [8]). As one can notice, the LSVM generally outperforms the SVM; in particular, considering the last two columns, an improvement of almost 3% is obtained using LSVM with $K = 7$ with respect to the linear SVM with $K = 1$ (single classifier).

In general, considering the last two columns, the AP gradually increases as the number of subcategories becomes larger, and it stabilizes around $K = 10$ and $K = 7$, for SVM and LSVM, respectively. This analysis suggests that the optimal value of K mostly depends on the data and not in the classification framework. Instead, the number of subcategories does affect the time performance because larger values of K correspond to higher testing time. For this reason, in order to find a good trade-off between testing time and detection accuracy, in the implementation adopted for the experiments presented in Sec. 3.2 we decided to set the number of subcategories equal to 7.

4. Conclusions

This paper presented a novel multi-cue pedestrian detection approach that encodes the high variability of un-occluded and occluded pedestrians by learning latent subcategory models. Combining this framework with a self-similarity based descriptor that encodes occlusion patterns over the depth silhouette, we set the state-of-the-art performance on the challenging Daimler Pedestrian Detection Benchmark Dataset [6]. Results show the effectiveness of our proposed system and its feasibility to be implemented in real-world systems. Since High Definition (HD) video surveillance cameras are becoming more and more popular, in the future work it would be valuable to extend ideas presented here to the setting of DPMs, in which the subclassifiers are part based models. Finally, we make our proposed system's code available for benchmarking purposes and to stimulate further research.

References

[1] R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? In *ECCV Workshops*, 2014.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[3] T. Deselaers and V. Ferrari. Global and efficient self-similarity for object classification and detection. In *CVPR*, 2013.

[4] S. K. Divvala, A. A. Efros, and M. Hebert. How important are deformable parts in the deformable parts model? In *ECCV Workshop*, 2012.

[5] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *PAMI*, 2011.

[6] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila. Multi-cue pedestrian classification with partial occlusion handling. In *CVPR*, 2010.

[7] M. Enzweiler and D. M. Gavrila. A multilevel Mixture-of-Experts framework for pedestrian classification. *TIP*, 2011.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge 2007 (voc2007).

[9] P. F. Felzenszwalb, R. Girshick, and D. McAllester. Discriminatively trained deformable part models - release 4.

[10] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Cascade object detection with deformable part models. In *CVPR*, 2010.

[11] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 32(9), 2010.

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

[13] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28, 1998.

[14] R. B. Girshick, P. F. Felzenszwalb, and D. A. McAllester. Object detection with grammar models. In *NIPS*, 2011.

[15] R. B. Girshick and J. Malik. Training deformable part models with decorrelated features. In *ICCV*, 2013.

[16] O. H. Jafari, D. Mitzel, and B. Leibe. Real-time rgb-d based people detection and tracking for mobile robots and head-worn cameras. In *ICRA*, 2014.

[17] S. Martelli, M. Cristani, L. Bazzani, D. Tosato, and V. Murino. Joining feature-based and similarity-based pattern description paradigms for object detection. In *ICPR*, 2012.

[18] F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *CVPR*, 2005.

[19] M. San Biagio, M. Crocco, M. Cristani, S. Martelli, and V. Murino. Heterogeneous auto-similarities of characteristics (hasc): exploiting relational information for classification. In *ICCV*, 2013.

[20] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake. Efficient human pose estimation from single depth images. *PAMI*, 35(12), 2013.

[21] S. Sridhar, A. Oulasvirta, and C. Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *ICCV*, 2013.

[22] M. J. Swain and D. H. Ballard. Color indexing. *IJCV*, 7(1), 1991.

[23] S. Tang, M. Andriluka, and B. Schiele. Detection and tracking of occluded people. *IJCV*, 110(1):58–69, 2014.

[24] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *ICCV*, 2003.

[25] S. Walk, K. Schindler, and B. Schiele. Disparity Statistics for Pedestrian Detection : Combining Appearance, Motion and Stereo. In *ECCV*, 2010.

[26] Y. Zheng, C. Shen, R. Hartley, and X. Huang. Effective pedestrian detection using center-symmetric local binary/trinary patterns. *arXiv preprint arXiv:1009.0892*, 2010.