Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection

Tomoki Watanabe, Satoshi Ito, and Kentaro Yokoi

Corporate Research and Development Center, TOSHIBA Corporation,
1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582, Japan
{tomoki8.watanabe,satoshi13.ito,kentaro.yokoi}@toshiba.co.jp

Abstract. The purpose of this paper is to detect pedestrians from images. This paper proposes a method for extracting feature descriptors consisting of co-occurrence histograms of oriented gradients (CoHOG). Including co-occurrence with various positional offsets, the feature descriptors can express complex shapes of objects with local and global distributions of gradient orientations. Our method is evaluated with a simple linear classifier on two famous pedestrian detection benchmark datasets: "*DaimlerChrysler pedestrian classification benchmark dataset*" and "*INRIA person data set*". The results show that proposed method reduces miss rate by half compared with HOG, and outperforms the state-of-the-art methods on both datasets.

Keywords: Pedestrian detection, CoHOG, co-occurrence histograms of oriented gradients, co-occurrence matrix.

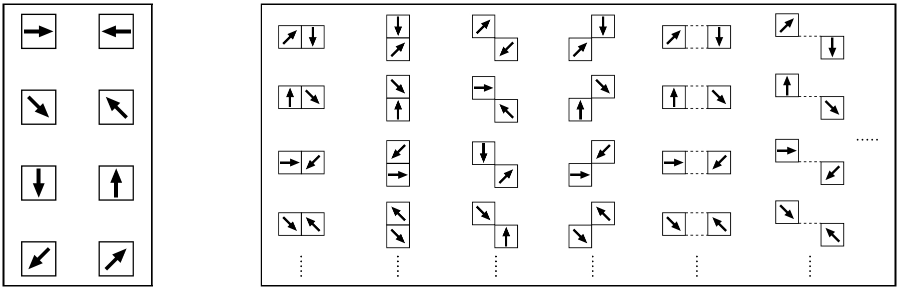
1 Introduction

Detecting pedestrians in images is essential in many applications such as automatic driver assistance, image surveillance, and image analysis. Extensive variety of postures and clothes of pedestrians makes this problem challenging.

Many types of feature descriptors have been proposed for pedestrian detection. Gavrila et al. used templates of pedestrian contours with chamfer matching [1], and LRF (Local Receptive Fields) with a quadratic SVM classifier [2]. They also combined those feature descriptors [3]. LRF are weight parameters of hidden layers of neural network which extract local feature of pedestrians. Viola et al. proposed a motion feature descriptor and combined it with cascaded AdaBoost classifier [4]. Papageorgiou et al. used SVM-based parts detectors with Haar wavelet feature and integrated them with SVM [5], [6].

Recently, using gradient-orientation-based feature descriptors, such as SIFT (Scale Invariant Feature Transform) [7] and HOG (Histograms of Oriented Gradients) [8], is a trend in object detection [9], [10]. Those feature descriptors are also used for pedestrian detection [8],[11],[12],[13]. Shashua et al. used body parts detectors using SIFT [11] and Mikolajczyk et al. also used jointed SIFT with an SVM classifier [12]. Dalal et al. proposed HOG and combined it with an SVM classifier [8], and also extended their method to motion feature descriptors [13].

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 37–47, 2009.



(a) Single

(b) Paired

Fig. 1. Vocabulary of gradient orientations. Though (a) a single gradient orientation has only eight varieties, (b) a pair of them has many more varieties than the single one.

Some multiple-edge-based feature descriptors also have been proposed. Wu et al. proposed edgelet feature descriptor which expresses long curves of edges [14]. Sabzmeydani et al. proposed shapelet feature descriptor based on selected edges by AdaBoost [15]. Since shapelets are the combination of edges, they can express more detailed shape information than what SIFT/HOG feature descriptors can.

We propose a multiple-gradient-orientation-based feature descriptor named "Co-occurrence Histograms of Oriented Gradients (CoHOG)". CoHOG is histograms of which a building block is a pair of gradient orientations. Since the pair of gradient orientations has more vocabulary than single one as shown in Fig. 1. CoHOG can express shapes in more detail than HOG, which uses single gradient orientation. Benchmark results on two famous datasets: DaimlerChrysler pedestrian classification benchmark dataset and INRIA person data set, show the effectiveness of our method.

The rest of this paper is organized as follows: Section 2 explains the outline of our pedestrian detection approach; Section 3 briefly explains HOG, and then describes our feature descriptor; Section 4 shows experimental results on two benchmark datasets; The final section is the conclusion.

2 Outline of Our Approach

In most pedestrian detection tasks, classification accuracy is the most important requirement. The performance of the system depends on the effectiveness of feature descriptors and the accuracy of classification models.

In this paper, we focus on the feature descriptor. An overview of our pedestrian detection processes is shown in Fig. 2. The first two parts extract feature descriptors from input images, and then the last part classifies and outputs classification results. We propose a high-dimensional feature descriptor in Section 3. Our feature descriptor is effective for classification, because it contains building blocks that have an extensive vocabulary.

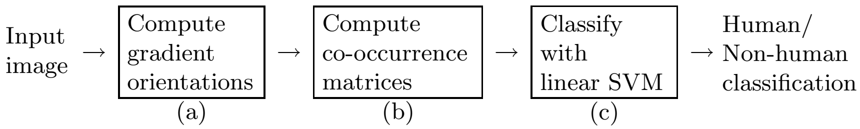

Input image
Compute gradient orientations
(a)
Compute co-occurrence matrices
(b)
Classify with linear SVM
(c)
Human/ Non-human classification

Fig. 2. Our classification process. We combine strong feature descriptor CoHOG and a conventional simple classifier. Our classification process consists of three parts: (a) computation of gradient orientations from input images, (b) computation of CoHOG from gradient orientations, and (c) classification with linear SVM classifier which is fast at learning and classification.

If the feature descriptor is informative enough, a simple linear classifier can detect pedestrians accurately. We use a linear classifier obtained by a linear SVM [16] which works fast at learning and classification.

3 Gradient Orientation Based Feature Descriptor

3.1 Histograms of Oriented Gradients (HOG)

We briefly explain the essence of the HOG calculation process with Fig. 3. In order to extract HOG from an image, firstly gradient orientations at every pixel are calculated (Fig. 3(a)). Secondly a histogram of each orientation in a small rectangular region is calculated (Fig. 3(b)). Finally the HOG feature vector is created by concatenating the histograms of all small regions (Fig. 3(c)).

HOG has two merits for pedestrian detection. One merit is the robustness against illumination variance because gradient orientations are computed from local intensity difference. The other merit is the robustness against deformations because slight shifts and affine deformations make small histogram value changes.

3.2 Co-occurrence Histograms of Oriented Gradients (CoHOG)

We propose a high-dimensional feature “Co-occurrence Histograms of Oriented Gradients (CoHOG)”. Our feature uses pairs of gradient orientations as units,

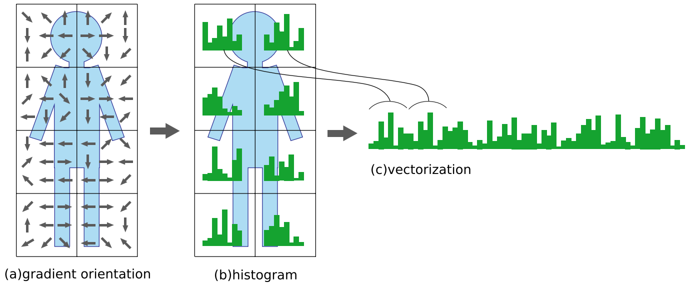


Fig. 3. Overview of HOG calculation

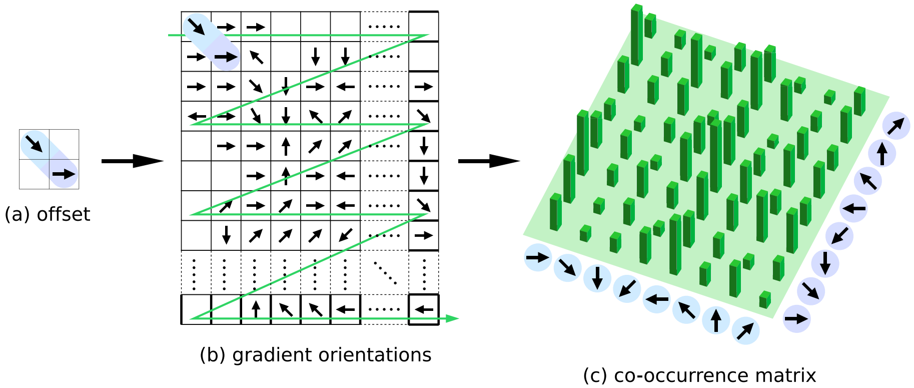


Fig. 4. Co-occurrence matrix of gradient orientations. It calculates sums of all pairs of gradient orientations at a given offset.

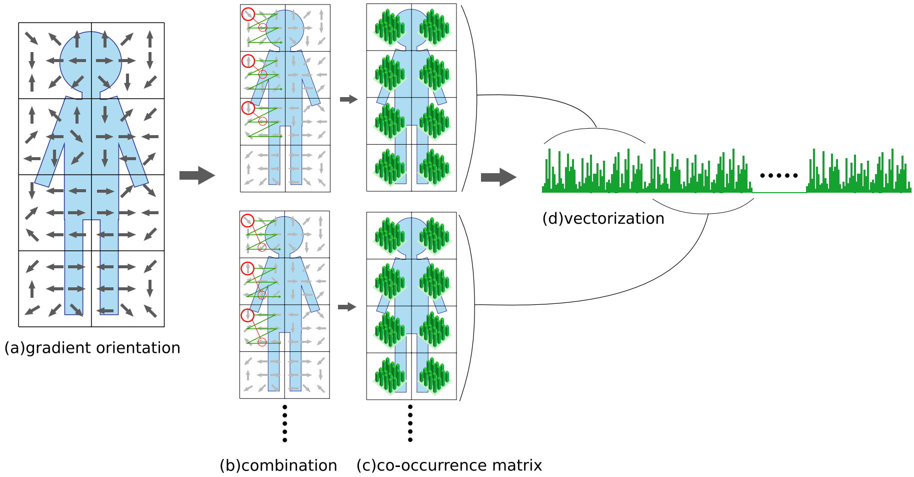


Fig. 5. Overview of CoHOG calculation

from which it builds the histograms. The histogram is referred to as the co-occurrence matrix, hereafter. The co-occurrence matrix expresses the distribution of gradient orientations at a given offset over an image as shown in Fig. 4. The combinations of neighbor gradient orientations can express shapes in detail. It is informative for pedestrian classification. Mathematically, a co-occurrence matrix C is defined over an $n \times m$ image I, parameterized by an offset (x, y), as:

$$C_{x,y}(i,j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, & \text{if } I(p,q) = i \text{ and } I(p+x, q+y) = j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

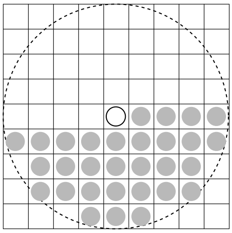


Fig. 6. Offsets of co-occurrence matrices. Offsets are smaller than the large dashed-circle. The center small white-circle and the other 30 dark-circles are paired. We calculate 31 Co-occurrence matrices with different offsets including zero offset.

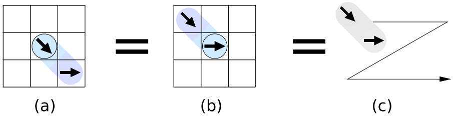


Fig. 7. Offset values of (a) $(1, 1)$ and (b) $(-1, -1)$ are different, but they behave as same as the other in the calculation of co-occurrence matrix

```
given $I$: an image of gradient orientation
initialize $H \leftarrow 0$
for all positions $(p, q)$ inside of the image do
    $i \leftarrow I(p, q)$
    $k \leftarrow$ the small region including $(p, q)$
    for all offsets $(x, y)$ such that corresponds neighbors do
        if $(p + x, q + y)$ is inside of the image then
            $j \leftarrow I(p + x, q + y)$
            $H(k, i, j, x, y) \leftarrow H(k, i, j, x, y) + 1$
        end if
    end for
end for
```

Fig. 8. Implementation of CoHOG calculation. The bins of histogram H are initialized to zero before voting. All pixels in the gradient orientation image I are scanned, and bins of H corresponding to pixels are incremented.

CoHOG has robustness against deformation and illumination variance for the same reasons as HOG, because CoHOG is gradient based histogram feature descriptor.

We describe the processes of CoHOG calculation shown in Fig. 5. Firstly, we compute gradient orientations from an image by

$$\theta = \arctan \frac{v}{h}, \tag{2}$$

where v and h are vertical and horizontal gradient respectively calculated by Sobel filter, Roberts filter, etc. We label each pixel with one of eight discrete orientations or as no-gradient (Fig. 5(a)). All $0° - 360°$ orientations are divided into eight orientations per 45°. No-gradient means $\sqrt{v^2 + h^2}$ is smaller than a threshold. Secondly, we compute co-occurrence matrices by Eq. (1) (Fig. 5(b)). The offsets we used are shown in Fig. 6. By using short-range and long-range offsets, the co-occurrence matrix can express local and global shapes. We do not use half of the offsets, because they behave as same as the others in calculation of co-occurrence matrix as shown in Fig. 7. The dashed-circle is the maximum range of offsets. We can get 31 offsets including a zero offset. The co-occurrence matrices are computed for each small region (Fig. 5(c)). The small rectangular regions are tiled $N \times M$, such as 3×6 or 6×12, with no overlapping. Finally, the components of all the co-occurrence matrices are concatenated into a vector (Fig. 5(d)).

Since CoHOG expresses shapes in detail, it is high-dimensional. The dimension is $34,704$, when the small regions are tiled 3×6. From one small region, CoHOG obtains 31 co-occurrence matrices. A co-occurrence matrix has 64 components (Fig. 4(c)). The co-occurrence matrix calculated with zero offset has only eight effective values because non-diagonal components are zero. Thus CoHOG obtains $(64 \times 30 + 8) \times (3 \times 6) = 34,704$ components from an image. In fact, the effective values are fewer than $34,704$, because co-occurrence matrices have multiple zero valued components. Zero valued components are not used in classification, because their inner product is zero at all times. Nevertheless, CoHOG is a more powerful feature descriptor than HOG.

The implementation of CoHOG is simple. An example of CoHOG implementation is shown in Fig. 8. We can calculate CoHOG by only iterating to increment the components of co-occurrence matrices, whereas HOG calculation includes more procedures, such as orientation weighted voting, histogram normalization, region overlapping, and etc. CoHOG can achieve high performance without those complex procedures.

4 Experimental Results

We evaluated the performance of CoHOG by applying our method to two pedestrian image datasets: the DiamlerChrysler dataset [2] and the INRIA dataset [8], which are widely used pedestrian detection benchmark datasets. The DaimlerChrysler dataset contains human images and non-human images cropped into 18×36 pixels. the INRIA dataset contains human images cropped 64×128 pixels and non-human images of various sizes. The details of those datasets are shown in Table 1, and some samples of the datasets are shown in Fig. 9.

Because the size of the images are different, in our method we divided the DiamlerChrysler dataset images into 3×6 small regions, and the INRIA dataset images into 6×12 small regions. Thus the dimension of our feature is $34,704$ on the DiamlerChrysler dataset, and quadruple that on the INRIA dataset. We used a linear SVM classifier trained with LIBLINEAR [17] which solves linear

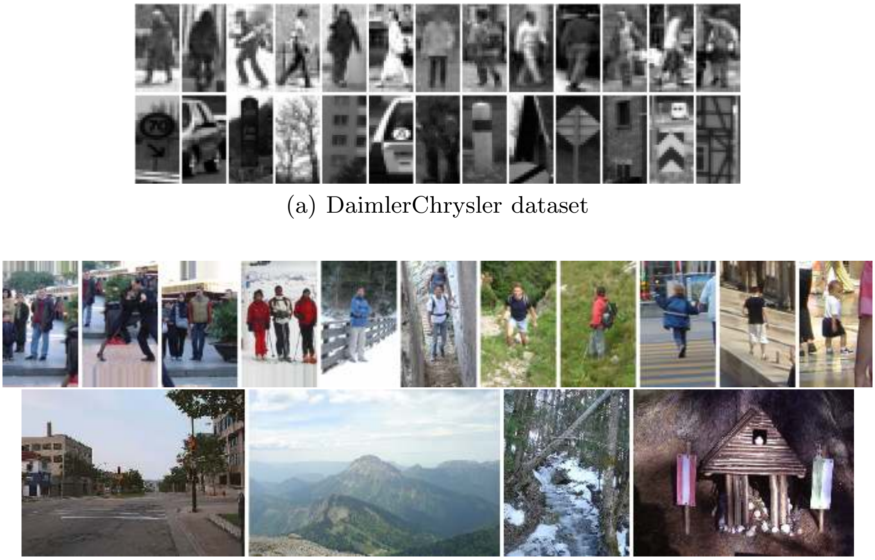
Table 1. Pedestrian detection benchmark datasets

(a) DaimlerChrysler dataset

Dataset Name	DaimlerChrysler Pedestrian Classification Benchmark Dataset
Distribution site	http://www.science.uva.nl/research/isla/downloads/pedestrians/
Training data	4,800 × 3 human images 5,000 × 3 non-human images
Test data	4,800 × 2 human images 5,000 × 2 non-human images
Image size	18 × 36 pixels

(b) INRIA dataset

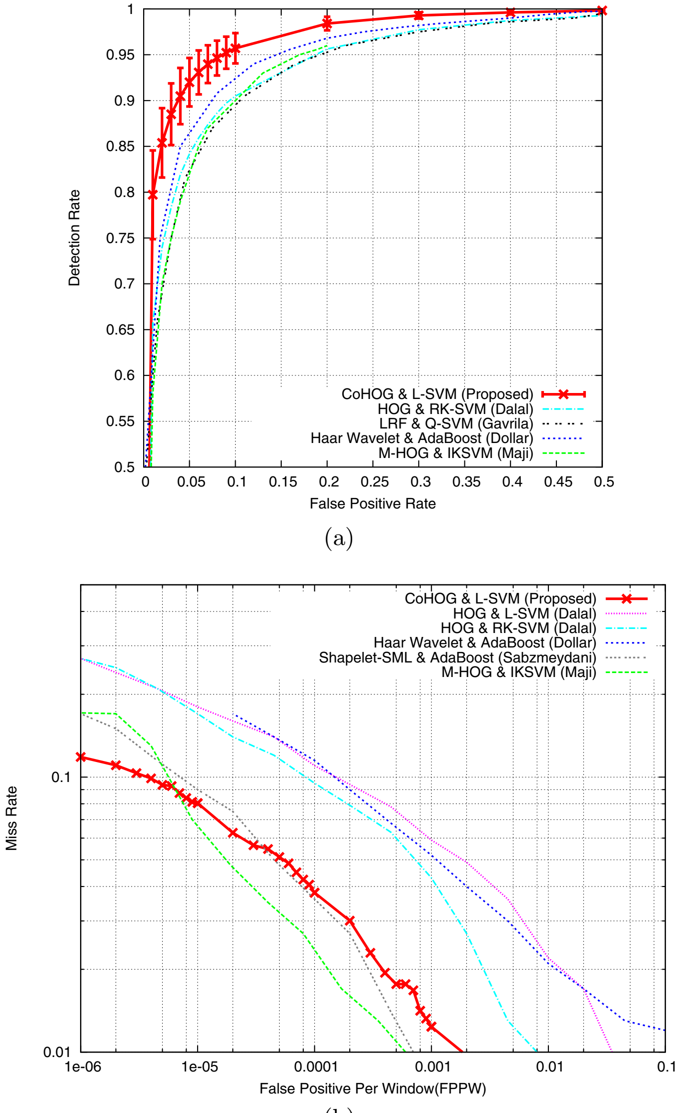
Dataset Name	INRIA Person Data Set
Distribution site	http://pascal.inrialpes.fr/data/human/
Training data	2,716 human images 1,218 non-human images (10 regions are randomly sampled per an image for training.)
Test data	1,132 human images 453 non-human images
Image size	Human images are 64 × 128 pixels Non-human images are various size (214 × 320 – 648 × 486 pixels)



(a) DaimlerChrysler dataset

(b) INRIA dataset

Fig. 9. Thumbnails of (a) DaimlerChrysler dataset and (b) INRIA dataset. Upper rows are images of humans and lower rows are images of non-humans in each dataset.



(a)

(b)

Fig. 10. Performance of our methods on (a) DaimlerChrysler dataset and (b) INRIA dataset. We compared our method with several previous methods. Our method shows the best performance on the DaimlerChrysler dataset. Miss rate improves more than 40% from that of the state-of-the-art method at a false positive rate of 0.05. On the INRIA dataset, our method decreased miss rate by 30% from that of the state-of-the-art method at a FPPW of 10^{-6}. Our method reduces miss rate by half compared with HOG on both datasets.

SVM learning problems much faster than previous solvers such as LIBSVM [18] and SVMLight [19].

We compared our method with five previous methods [8], [2], [20], [15], [21]. All the methods use different features and classifiers: Dalal et al. used HOG, and RBF kernel SVM and linear SVM [8]; Gavrila et al. used local receptive fields (LRF) and quadratic SVM [2]; Dollar et al. used Haar wavelet and AdaBoost [20]; Sabzmeydani et al. used shapelet and AdaBoost [15]; and Maji et al. used multi-level oriented edge energy features and intersection kernel SVM (IKSVM) [21].

The comparison of their performances is shown in Fig. 10. The results of previous methods are traced from the original papers except the performance of HOG on the DaimlerChrysler dataset, because it is not shown by Dalal et al. We show it based on the result of our experiment. The parameters of HOG are as follows: Nine gradient orientations in 0°–180°, cell size of 3×3 pixels, block size of 2×2 cells, L2Hys normalized. The classifier is an RBF-kernel SVM. In Fig. 10(a), ROC (Receiver Operating Characteristic) curves on the DaimlerChrysler dataset are shown. An ROC curve further towards the top-left of the diagram means better performance. The results show that our method achieved the best detection rate at every false positive rate. Our method reduced the miss rate ($= 1 -$ detection rate) by about 40% from the state-of-the-art method at a false positive rate of 0.05; the miss rate of our method is 0.08 and that of Dollar et al., the second best, is 0.14.

In Fig. 10(b), DET (Detection Error Tradeoff) curves on the INRIA dataset are shown. A DET curve further towards the bottom-left of the diagram means better performance. The results show that the performance of our method is the best at low FPPW (False Positive Per Window) and comparable to the state-of-the-art method at other FPPW. Our method reduced miss rate by about 30% from the state-of-the-art method at a FPPW of 10^{-6}; the miss rate of of our method is 0.12 and the that of Maji et al. is 0.17. The performance at low FPPW is important for pedestrian detection, because most of the pedestrian detection systems work at low FPPW to improve usability with few false positives.

The results show that our method is better than the state-of-the-art methods or at least comparable. Furthermore, they show the stability of our method; the performance of the method of Dollar et al. is not good on the INRIA dataset and the method of Maji et al. is not good on the DaimlerChrysler dataset, however, the performance of our method is consistently good on both datasets. Though our method uses a linear classifier which is simpler than an RBF-kernel SVM classifier used with HOG, the miss rate of our method is less than half that of HOG.

5 Conclusion

In this paper, we proposed a high-dimensional feature descriptor "Co-occurrence histograms of oriented gradients (CoHOG)" for pedestrian detection. Our feature descriptor uses pairs of gradient orientations as units, from which it builds histograms. Since the building blocks have an extensive vocabulary, our feature descriptor can express local and global shapes in detail. We compared the classification performance of our method and several previous methods on two famous datasets. The experimental results show that the performance of our method is better than that of the state-of-the-art methods or at least comparable, and consistently good on both datasets. The miss rate (i.e. the rate of human images classified as non-human) of our method is less than half that of HOG. Future work involves applying the proposed feature descriptor to other applications.

References

1. Gavrila, D., Philomin, V.: Real-time object detection for “smart” vehicles. In: The Seventh IEEE International Conference on Computer Vision, vol. 1, pp. 87–93. IEEE Computer Society Press, Los Alamitos (1999)
2. Munder, S., Gavrila, D.M.: An experimental study on pedestrian classification. IEEE Trans. Pattern Anal. Mach. Intell. 28(11), 1863–1868 (2006)
3. Gavrila, D.M., Munder, S.: Multi-cue pedestrian detection and tracking from a moving vehicle. Int. J. Comput. Vision 73(1), 41–59 (2007)
4. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: The Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, pp. 734–741. IEEE Computer Society, Los Alamitos (2003)
5. Mohan, A., Papageorgiou, C., Poggio, T.: Example-based object detection in images by components. IEEE Trans. Pattern Anal. Mach. Intell. 23(4), 349–361 (2001)
6. Papageorgiou, C., Poggio, T.: A trainable system for object detection. Int. J. Comput. Vision 38(1), 15–33 (2000)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 257–263 (2003)
10. Winder, S.A.J., Brown, M.: Learning local image descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
11. Shashua, A., Gdalyahu, Y., Hayun, G.: Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In: IEEE Intelligent Vehicles Symposium, pp. 1–6 (2004)
12. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 69–82. Springer, Heidelberg (2004)
13. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006)
14. Wu, B., Nevatia, R.: Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In: The Tenth IEEE International Conference on Computer Vision, Washington, DC, USA, vol. 1, pp. 90–97. IEEE Computer Society Press, Los Alamitos (2005)

15. Sabzmeydani, P., Mori, G.: Detecting pedestrians by learning shapelet features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
16. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20(3), 273–297 (1995)
17. Hsieh, C., Chang, K., Lin, C., Keerthi, S., Sundararajan, S.: A dual coordinate descent method for large-scale linear svm. In: McCallum, A., Roweis, S. (eds.) The 25th Annual International Conference on Machine Learning, pp. 408–415. Omnipress (2008)
18. Hsu, C.W., Chang, C.C., Lin, C.J.: A practical guide to support vector classification. Technical report, Taipei (2003)
19. Joachims, T.: Training linear svms in linear time. In: The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 217–226 (2006)
20. Dollar, P., Tu, Z., Tao, H., Belongie, S.: Feature mining for image classification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
21. Maji, S., Berg, A.C., Malik, J.: Classification using intersection kernel support vector machines is efficient. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2008)