2010

Human detection using local shape and non-redundant binary patterns

Duc Thanh Nguyen

University of Wollongong, dtn156@uow.edu.au

Wanqing Li

University of Wollongong, wanqing@uow.edu.au

Philip Ogunbona

University of Wollongong, philipo@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>

Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Nguyen, Duc Thanh; Li, Wanqing; and Ogunbona, Philip: Human detection using local shape and non-redundant binary patterns 2010, 1145-1150.
<https://ro.uow.edu.au/infopapers/2188>

Human detection using local shape and non-redundant binary patterns

Abstract

Motivated by the advantages of using shape matching technique in detecting objects in various postures and viewpoints and the discriminative power of local patterns in object recognition, this paper proposes a human detection method combining both shape and appearance cues. In particular, local shapes of the body parts are detected using template matching. Based on body parts' shapes, local appearance features are extracted. We introduce a novel local binary pattern (LBP) descriptor, called Non-Redundant LBP (NRLBP), to encode local appearance of human. The proposed method was evaluated and compared with other state-of-the-art human detection methods on two commonly used datasets: MIT and INRIA pedestrian test sets. We also performed extensive experiments on selecting appropriate parameters as well as verifying the improvement of the proposed method through all stages of the framework.

Keywords

shape, non, redundant, binary, patterns, local, detection, human

Disciplines

Physical Sciences and Mathematics

Publication Details

Nguyen, D., Li, W. & Ogunbona, P. (2010). Human detection using local shape and non-redundant binary patterns. International Conference on Control, Automation, Robotics and Vision (pp. 1145-1150). Piscataway, New Jersey, USA: IEEE.

Human Detection Using Local Shape and Non-Redundant Binary Patterns

Duc Thanh Nguyen, Wanqing Li, and Philip Ogunbona
Advanced Multimedia Research Lab, ICT Research Institute
School of Computer Science and Software Engineering
University of Wollongong, Australia

Abstract—Motivated by the advantages of using shape matching technique in detecting objects in various postures and viewpoints and the discriminative power of local patterns in object recognition, this paper proposes a human detection method combining both shape and appearance cues. In particular, local shapes of the body parts are detected using template matching. Based on body parts' shapes, local appearance features are extracted. We introduce a novel local binary pattern (LBP) descriptor, called Non-Redundant LBP (NRLBP), to encode local appearance of human. The proposed method was evaluated and compared with other state-of-the-art human detection methods on two commonly used datasets: MIT and INRIA pedestrian test sets. We also performed extensive experiments on selecting appropriate parameters as well as verifying the improvement of the proposed method through all stages of the framework.

Index Terms—Human detection, local binary patterns.

I. INTRODUCTION

Human detection from images is an active research topic in computer vision. The challenge of the task arises from the numerous variations that human postures can assume and the complexity the surrounding environment can be (e.g. cluttered background, crowded scene, etc.).

A number of approaches have been proposed in the literature [1]. Generally speaking, existing human detection methods can be categorized into template matching based approach and learning based approach. In the template matching based approach, humans are described explicitly by full body [2], [3] or body part templates [4], [5] and the task of human detection becomes to find the best matching templates given an input image. The templates can be represented as intensity/color images [5] when the appearance of the humans is considered or simply as binary contours [2], [3], [4] when shape information is employed. Template matching based approach has several advantages. First, human's shape can be well captured by the contour templates. In addition, this approach allows the variation of human postures and viewpoints. However, the contour templates are required to be given in advance.

In the learning based approach, appearance and shape features, used to describe humans are obtained through training classifiers such as SVM, AdaBoost, etc. The problem is then often formulated as the binary classification. Examples of this approach are [6], [7], [8], [9], [10], [11], [12], [13]. Since the spatial information of body parts is encoded and included in the features, these methods limit the variation of human postures and viewpoints to what often occur in the training dataset.

Motivated by the advantages of using contour templates in detecting object's shape at various postures and viewpoints, and the discriminative ability of local patterns in object recognition, this paper introduces a human detection method with the following contributions. First, we present a human detection framework integrating multiple cues including local shape and appearance as follows. Shapes of the body parts are determined using an improved template matching method that combines both the strength and orientation of edge pixels. Based on the body parts' shapes represented by contours, local appearance patterns are extracted along the matching contours. The second contribution of this paper is to propose a novel LBP, called Non-Redundant LBP (NRLBP), to encode local appearance of human. Experiments have verified the robustness of the proposed method in detecting humans in multiple views and postures and under cluttered backgrounds.

The rest of the paper is organized as follows. In section II, we briefly review the related works. Section III presents the appearance features (the novel LBP). The proposed human detection method is presented in section IV. Experimental results along with some comparative analysis are shown in Section V. Section VI concludes the paper with remarks.

II. RELATED WORK

Template matching based approach often requires templates as two-dimensional contours representing humans in various postures and viewpoints. For example, Gavrilu et al. [2], [3] clustered full body human templates into a hierarchical structure where the dissimilarity between two templates was defined by the Chamfer distance. Since the above works focus on detection of a full human body, we refer to them as global detection methods.

Alternatively, local methods detect humans by locating the body parts. For example, Lin et al. [4] decomposed a human body structure into a hierarchical tree of body parts including head-torso, upper legs, and lower legs. The detection was then performed by detecting individual body parts sequentially in the hierarchical tree.

Some methods combine both local and global detection [5]. For example, Leibe et al. [5] used a codebook of local shapes to describe the human body structure. The relationship between local and global shapes was learned through training. The local shapes were further employed to vote for the global shapes and the Chamfer matching was then applied to select the best fit

of the joint global and local detection responses.

Learning based human detection algorithms often proceed by learning shape and appearance features from training data. This approach does not require templates but shape and appearance features are defined through training human/non-human patterns. For example, Mohan et al. [6] used Haar wavelets to describe body parts while Wu et al. [8] introduced a so-called "edgelet" feature. Viola et al. [7] employed the rectangular features [14] computed on the difference images to encode motion patterns. In [9], a well-known feature, called histogram of oriented gradients (HOG) was introduced. This feature then has been received much attention and considered as state-of-the-art feature for object detection. In [10], covariance based features were employed. Recently, local binary pattern (LBP), originally proposed for texture classification [15], was used to encode human appearance [12].

Another aspect of learning based human detection is to develop robust learning algorithms. For example, a cascade method proposed in [7] was employed for fast human detection in [16]. In [11], a meta-stage was added to the cascade Adaboost to exploit the inter-stage information. Employing various features [11], [13] could provide richer descriptors but might lead learning algorithms such as SVMs to be intractable with respect to training. This problem was addressed in [17] and Partial Least Squares (PLS) analysis, a dimensionality reduction technique, was used. In [18], histogram intersection kernel SVM was introduced for fast training and classification. A study on features and classifiers selection for human detection was conducted recently by Wojek et al. in [19].

III. NON-REDUNDANT LBP (NRLBP)

A. LBP

Original LBP was developed for texture classification [15]. The advantages of LBP are its robustness under illumination changes, computational simplicity and discriminative power. We adopt the notation of LBP as follows. Given a pixel $c = (x_c, y_c)$, the value of the LBP code of c is given by:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (1)$$

where P is the number of sampled points (neighbor pixels of c) whose the distances to c do not exceed the radius R , g_c and g_p are the intensities of c and p (a neighbor pixel of c) respectively, and

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Fig 1 represents an example of the LBP, called original LBP hereafter, in which the LBP code of the center pixel (in red color) is obtained by comparing its intensity with neighbor pixels' intensities. The neighbor pixels whose intensities are equal or higher than the center one's are labeled as "1", otherwise as "0". In this example, since we consider all 8-neighbor pixels of the center pixel and the distance between a neighbor pixel to the center pixel is computed using l_∞ , R

Fig. 1. An illustration of the original LBP descriptor.

Fig. 2. The left and right image represents the same human and in the same scene but with different contrasts between the background and foreground. Notice that the LBP code at (x, y) (in the left and right image) is complementary each other, i.e. the sum of these codes is $2^P - 1$.

and P are 1 and 8 respectively. The followings are important properties of the LBP descriptor.

Uniform and non-uniform LBP. Uniform LBP is defined as the LBP that has at most two bitwise transitions from 0 to 1 and vice versa in its circular binary representation. LBPs which are not uniform are called non-uniform LBPs. As indicated in [15], an important property of uniform LBPs is the fact that they often represent primitive structures of the texture while non-uniform LBPs usually correspond to unexpected noises and thus they are less discriminative.

LBP histogram. Scanning a given image in pixel by pixel fashion, LBP codes are accumulated into a discrete histogram called LBP histogram. Intuitively, the number of $LBP_{P,R}$ histogram bins is 2^P .

B. Non-Redundant LBP

Although the robustness of the original LBP has been proved in many applications, it has drawbacks when employed to encode human appearance. The disadvantages can be considered with respect to two aspects:

- **Storage ability:** as presented, the original LBP requires 2^P bins of histogram.
- **Discriminative ability:** the original LBP is sensitive to the relative changes between the background and foreground (the region inside the human body). It rigorously depends on the intensities of particular locations and thus varies based on the human's clothing which is often varied. For example, the LBP codes of the regions indicated by red rectangles (small box) in both the left and right images in Fig. 2 are quite different while they actually represent the same structure.

To overcome both of the above issues, we propose a novel LBP, called Non-Redundant LBP (NRLBP), as follows,

$$\begin{aligned} NRLBP_{P,R}(x_c, y_c) \\ = \min \left\{ LBP_{P,R}(x_c, y_c), 2^P - 1 - LBP_{P,R}(x_c, y_c) \right\} \end{aligned} \quad (2)$$

Intuitively, the NRLBP considers a LBP code and its complement as same. For example, the two LBP codes "11000011" and "00111100" in Fig. 2 are counted once. Obviously, with the NRLBP, the number of bins in the LBP histogram is reduced to the half. Furthermore, compared with the original LBP, the NRLBP is more discriminative in this case since it reflects the relative contrast between the background and foreground rather than forces a fixed arrangement of the background and foreground. Hence, the NRLBP is more robust and adaptive with changes of the background and foreground. The robustness of the NRLBP is also verified in experiments (see section V).

IV. PROPOSED HUMAN DETECTION METHOD

The proposed human detection method can be summarized as follows. First, given a detection window, shapes of the body parts are determined using template matching. The outcome of this stage is a set of partial contours describing shapes of the body parts. After that, local appearance features defined as the histograms of NRLBPs are computed along the detected partial contours. Finally, a feature vector is formed by concatenating the matching costs of individual pixels, individual parts, called shape feature, and NRLBP histograms. Feature vectors corresponding to the positive (human) and negative (non-human) samples are collected to train a SVM classifier. The proposed human detection framework is presented in Fig. 3.

A. Shape Matching

We employ a set of contour templates including top (head-torso), bottom (legs), left, and right parts to model a full human body structure (see Fig. 4). To create these part templates, a number of templates representing full human body shape are first collected. Examples of these templates are shown in Fig. 4(a). Each template is centered in a 30×60 window and then divided into 4 parts: top, bottom, left, and right. For each type of parts, e.g. leg, the part templates are clustered using a K -means algorithm and the mean templates are considered as the prototype part templates. Let $\mathcal{P}_i$ be the sets of prototypes for $i = top, bottom, left, right$ part respectively. In our implementation, $|\mathcal{P}_{top}| = 5, |\mathcal{P}_{bottom}| = 8, |\mathcal{P}_{left}| = |\mathcal{P}_{right}| = 6$. Fig. 4 (c, d, e, f) shows the prototypes.

Given a detection window W , using the template matching method proposed in [20], the best matching human posture (configuration), i.e. a set of best matching templates $C^* = \{T_i^*\}, i = top, bottom, left, right$ can be determined as follows.

$$T_i^* = \arg \min_{T \in \mathcal{P}_i} D(T, W_i) \quad (3)$$

where W_i is a subpart of the detection window W corresponding to the top, bottom, left, or right part respectively. $D(T, W_i)$

Fig. 4. Part templates used in the proposed method.

is the dissimilarity (matching cost) between the template T and image region in W_i which is computed as,

$$D(T, W_i) = \sum_{t \in T} \omega_T(t) d_{T, W_i}(t) \quad (4)$$

where $d_{T, W_i}(t)$ is the spatial-oriented distance between a point $t \in T$ and its closest edge point in the edge image of W_i . $d_{T, W_i}(t)$ can be computed efficiently using a Generalized Distance Transform (readers are referred to [20] for more details).

The weight $\omega_T(t)$ in (4) represents the importance of a point $t \in T$. Different points should contribute differently in the matching cost $D(T, W_i)$. For instance, on the human body contour, the points along to the two curves of the head-shoulder template always appear in every human pattern thus more discriminative compared with the points belonging to the arms, which are varied and dependent on the human postures. In this paper, the weight $\omega_T(t)$ of a point t is calculated simply based on the frequency that t appears given a set of the templates of a same body part type, e.g. $\mathcal{P}_{left}$. In particular, for each $\mathcal{P}_i$ and for each template $T \in \mathcal{P}_i$, $\omega_T(t)$ can be computed as,

$$\omega_T(t) = \frac{\sum_{T' \in \mathcal{P}_i} (1 - d_{T, T'}(t))}{\sum_{t' \in T'} \sum_{T' \in \mathcal{P}_i} (1 - d_{T, T'}(t'))} \quad (5)$$

As can be seen that, for each detection window, the number of templates matched is $5 + 8 + 6 + 6 = 25$ (templates) to cover up to $5 \times 8 \times 6 \times 6 = 1440$ postures. Compared with full body detection approach, this is an advantage since the matching is performed on a small set of templates to cover a huge number of human postures.

B. Feature Extraction

It is reasonable to use the NRLBP to encode local appearance of human since the intensities of the foreground might be

Fig. 3. The proposed human detection framework.

either higher or lower than the intensities of the background (Fig. 2). In addition, the texture inside human body varies and depends on human's clothing. Therefore, the NRLBPs should be defined locally along the body contours to encode the contrast between the foreground and background rather than the textured structures inside human body. In this paper, local appearance features are computed as follows.

Once the best matching configuration C^* has been identified, for every template $T_i^* \in C^*$, a set of the closest points (on the edge image of W_i) can be determined. Since the number of points on the templates is different from one to another (even the templates represent the same type of body part), we uniformly sample the top, bottom, left, and right templates by 25, 15, 20, and 20 points respectively. Thus, the total number of sampled points for each posture (configuration) is 80.

For each sampled point on the template, its closest edge point on the edge image of W is determined and the NRLBP histogram of a $(2L+1) \times (2L+1)$ -local window W_L centered at this edge point is computed. Since we use the NRLBP, we could reduce the number of bins in the histogram from 59 (as the original LBP) to 30 in which all non-uniform NRLBPs vote for one bin and each uniform NRLBP is cast into a unique bin corresponding to its NRLBP code. The NRLBP histogram calculated from a local window W_L is then normalized using a normalization method, e.g. L_2 -norm, L_1 -norm, or L_1 -square norm. For fast computation of the NRLBP histogram, we employ the concept of "integral image" as proposed in [7]. Since the number of bins is only 30, it is quite possible to implement and store the integral image in computer memory. By computing the NRLBP histogram for every sampled point, we create a $80 \times 30 = 2400$ dimensional vector for each detection window W to describe local appearance of human. To encode shape information, the feature vector can be extended by combining with the matching costs of 80 sampled points, i.e. $d_{T,W_i}(\cdot)$, and the matching costs of 4 parts, i.e. $D(T, W_i)$ with $i = top, bottom, left, right$. Notice that the feature vector may need the matching costs of 4 parts since these costs might be obtained from more than 80 sampled points. Finally, we can create a rich feature vector with $2400 + 80 + 4 = 2484$ elements for each detection window W (see Fig. 5). The feature vectors of positive and negative samples are then used to train a SVM for classification.

Fig. 5. Shape-appearance feature vector. Left: input image, right: edges with the Generalized Distance Transformed image generated using [20].

V. EXPERIMENTAL RESULTS

A. Datasets

The proposed human detection method was evaluated on two popular pedestrian test sets: MIT dataset [6] and INRIA dataset [9]. The MIT dataset includes 924 positive samples containing pedestrians in frontal or back-view and no negative images. Compared with the MIT dataset, the INRIA dataset is more challenging in which pedestrians are in multiple views and extensive postures, and cluttered backgrounds with various illumination changes (see Fig. 7). The INRIA dataset contains 2416 positive training samples and 1218 negative training images plus 1126 positive testing samples and 453 negative testing images. For performance evaluation, we employed the DET (detection error trade-off) curves on the log-log scale to measure the miss rate vs. FPPW (false positive per window). In the followings, we present experiments on parameters and features selection, and comparisons with other state-of-the-art human detection methods.

B. Parameters and Features Selection

There are a number of parameters used in the proposed method including the number of sampled points (P) and radius (R) in the definition of the NRLBP (2), and local window's size (controlled by L). In addition, normalization methods and classifiers also affect the detection performance. To verify the effects of parameters, we conducted all experiments on the same dataset. The INRIA dataset was selected for this purpose. Furthermore, once a parameter is investigated, other parameters are maintained unchanged, i.e. at one time, only one parameter is changed. Finally, to avoid the explosion of all possible combinations of parameters, we assume that all parameters are

independent of each other. This means that we can investigate all parameters individually and the later experiment is based on the observation of the former experiment.

Based on the work of Wang et al. [13], we employed the $LBP_{8,1}$ for all experiments. However, we investigated the effects of various values of $L \in \{4, 5, 6, 7, 8\}$. The numerical data is provided in Table I. As shown in this table, $L = 7$, i.e. 15×15 local window, gives the best performance at low FPPW rates. However, the results corresponding to other values of L such as 4, 5, 6, and 8 are slightly different. This means that the detection performance is not sensitive with the changes of the local window's size. Notice that, in this experiment, we only used the appearance, i.e. each feature vector has 2400 elements and L_1 -square norm was used for normalization of NRLBP histograms. In addition, we used a SVM for training and testing various parameters and features. The training and testing procedures were implemented similarly to [9]. The training dataset consists of 2416 positive samples and 12180 negative samples (created by selecting randomly 10 samples per negative image). The training step was performed only one time since all optimal parameters need to be determined before the second time of training.

TABLE I
FPPW (FIRST COLUMN) AND APPROXIMATED MISS RATES
CORRESPONDING TO VARIOUS VALUES OF L .

FPPW	$L = 4$	$L = 5$	$L = 6$	$L = 7$	$L = 8$
7×10^{-2}	0.0008	0.0017	0.0017	0.0000	0.0000
3×10^{-2}	0.0044	0.0026	0.0035	0.0026	0.0035
1×10^{-2}	0.0168	0.0142	0.0071	0.0079	0.0071
7×10^{-3}	0.0266	0.0230	0.0168	0.0097	0.0106
3×10^{-3}	0.0506	0.0435	0.0399	0.0364	0.0381
1×10^{-3}	0.0905	0.0923	0.0781	0.0692	0.0861
7×10^{-4}	0.1261	0.1012	0.0985	0.0896	0.1021
3×10^{-4}	0.1571	0.1527	0.1403	0.1367	0.1571

We also tried with different normalization methods and found that L_1 -square norm gives the best performance. This observation is consistent with the conclusion of Wang et al. [13]. For example, with $L = 7$ and at a FPPW rate of 10^{-3} , the miss rate corresponding to using L_1 -square norm was ≈ 0.0692 while this number was ≈ 0.0746 (increasing $\approx 0.5\%$) when using L_2 -norm.

For selecting classifiers, with $L = 7$ and at a FPPW rate of 10^{-3} , the linear SVM was incurred ≈ 0.1092 of miss rate, $\approx 1\%$ higher than the miss rate when using the kernel SVM. The difference then became significant ($\approx 6.75\%$) at a FPPW rate of 3×10^{-4} . Employing advanced classifiers, e.g. intersection kernel SVM [18], might improve the detection performance. However, development of robust classifiers is not the main focus of this paper.

We also compared the performance of using only appearance (NRLBP or original LBP) and combining both appearance and shape information (see Fig. 6). Recall that, to encode appearance, the feature vector is created by concatenating the NRLBP histograms of all local windows W_L centered

Fig. 6. Comparison between using appearance features only and combining shape and appearance features (this figure is best viewed in color).

along the contour points. To encode shape information, the feature vector is extended by attaching the matching costs of individual contour points and the matching costs of the four parts. As shown in Fig. 6, combining shape and appearance features could improve the detection performance. In addition, the NRLBP outperforms the original LBP in term of accuracy. Furthermore, with the NRLBP we could reduce the dimensionality of the feature vectors, thus save the memory and speed up the detection task. Notice that only one round of training was applied for this experiment.

C. Performance Evaluation and Comparison

On the MIT dataset, we used 724 positive samples for training and 200 remaining samples for testing. Since the MIT dataset does not include negative samples, we used the same negative samples from the INRIA dataset similarly to [9]. The performance was evaluated on the 200 remaining positive samples and 1500 negative samples. At the miss rate of 0.5%, i.e. 99.5% of true detection, the proposed method achieved 0.00% of FPPW. This result is comparable with the result of [9] and better than that of [8].

On the INRIA dataset, once the optimal parameters and features have been determined, a second time of training was performed. We exhaustively scanned the negative images to find hard negative samples whose the confidence score (positive probability) is higher than a predefined threshold (0.3 in our experiment). We then created a new negative samples set with 35180 hard negative samples. This set and the set of original positive samples were used to train a kernel SVM once again. The DET curve of the proposed method after the second training is shown in Fig. 8.

In addition to conducting experiments on parameters and features selection, we also compared the proposed method with other state-of-the-art human detection methods. Since the main aim of this paper is to propose a new feature for human detection combining shape and appearance information, it is reasonable to compare the proposed detector with other state-of-the-art detectors with respect to feature aspect. In particular, in this paper we compared the proposed method with the work of Viola et al. [14] (rectangular feature), Mohan et al. [6]

Fig. 7. Detection results on the INRIA dataset. (a) True detection in which humans are in various viewpoints and articulations. (b) Miss detection. The number under the each image is its corresponding confidence score and the rejected margin is set to 0.5.

Fig. 8. Comparison between the proposed detector and other state-of-the-art detectors (this figure is best viewed in color).

(Haar feature), Dalal et al. [9] (HOG feature), Tuzel et al. [10] (covariance feature), and Mu et al. [12] (semantic LBP). Fig. 8 shows the DET curves of the proposed method and other methods. Some successful and missed detection results are represented in Fig. 7. It can be seen from Fig. 7 that the proposed method could detect humans in unusual postures and various views.

VI. CONCLUSIONS

This paper presents a human detection method combining both local shape and appearance features. Shape of the human's parts are captured by using template matching technique. Based on parts' shapes, local appearance patterns are extracted. In this paper, we propose a Non-Redundant LBP (NRLBP) descriptor to encode local appearance of human body. Experimental results show that the proposed method is able to detect humans in significant pose articulations and multiple views. Employing a cascade strategy for solving the occlusion problem would be our future work.

REFERENCES

- [1] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *PAMI*, vol. 31, no. 12, pp. 2179–2195, 2009.
- [2] D. M. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *ICCV*, 1999.
- [3] D. M. Gavrila, "A Bayesian, exemplar-based approach to hierarchical shape matching," *PAMI*, vol. 29, no. 8, pp. 1408–1421, 2007.
- [4] Z. Lin, L. S. Davis, D. Doermann, and D. DeMenthon, "Hierarchical part-template matching for human detection and segmentation," in *ICCV*, 2007.
- [5] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *CVPR*, 2005.
- [6] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *PAMI*, vol. 23, no. 4, pp. 349–361, 2001.
- [7] P. A. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *IJCV*, vol. 63, no. 2, pp. 153–161, 2005.
- [8] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *ICCV*, 2005.
- [9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005.
- [10] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *CVPR*, 2007.
- [11] Y. T. Chen and C. S. Chen, "A cascade of feed-forward classifiers for fast pedestrian detection," in *ACCV*, 2007.
- [12] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *CVPR*, 2008.
- [13] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *ICCV*, 2009.
- [14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, vol. 1, 2001, pp. 511–518.
- [15] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [16] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," in *CVPR*, 2006.
- [17] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *ICCV*, 2009.
- [18] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *CVPR*, 2008.
- [19] C. Wojek, S. Walk, and B. Schiele, "Multi-cue onboard pedestrian detection," in *CVPR*, 2009.
- [20] D. T. Nguyen, W. Li, and P. Ogunbona, "An improved template matching method for object detection," in *ACCV*, 2009.