

HUMAN DETECTION USING DISCRIMINATIVE AND ROBUST LOCAL BINARY PATTERN

Amit Satpathy^{1,2}

Xudong Jiang¹

How-Lung Eng²

¹Nanyang Technological University
School Of EEE
Nanyang Avenue, Singapore 639798
{amit0010,exdjiang}@ntu.edu.sg

²Institute for Infocomm Research
Agency for Science, Technology and Research
1 Fusionopolis Way, Connexis, Singapore 138632
hleng@i2r.a-star.edu.sg

ABSTRACT

Despite superior performance of Local Binary Pattern (LBP) in texture classification and face detection, its performance in human detection has been limited for two reasons. Firstly, LBP differentiates a bright human from a dark background and vice-versa. This increases the intra-class variation of humans. Secondly, LBP is contrast and illumination invariant. It does not discriminate between weak contrast local regions and similar strong contrast ones, resulting in a similar feature representation. Non-Redundant LBP (NRLBP) has been proposed to solve the first issue of LBP. However, an inherent limitation of NRLBP is that LBP codes and their complements in the same block are mapped to the same code. Furthermore, NRLBP, like LBP, is also contrast and illumination invariant. In this paper, we propose a novel edge-texture feature, Discriminative Robust Local Binary Pattern (DRLBP), for human detection. DRLBP alleviates the problems of LBP and NRLBP by considering the weighted sum and absolute difference of a LBP code and its complement. Our experimental results show that DRLBP consistently outperforms LBP and NRLBP for human detection.

Index Terms— local binary pattern, feature extraction, human detection, LBP, pedestrian detection

1. INTRODUCTION

A challenging problem of human detection to date is finding the most suitable feature for human description. The work on feature extraction can be divided into two groups based on the representation. The first group is *sparse* representation, where local features are obtained using interest point detection algorithms and are used to represent humans holistically (whole) or by parts. In [1, 2], parts-based models for detecting humans were proposed whereby features were extracted for each part and based on geometric constraints, assembled to form descriptors. Contour fragments of objects were used in [3] as features which were learnt incrementally and shared across object categories. Nguyen *et al.* [4] proposed using Non-Redundant Local Binary Pattern (NRLBP) to describe regions around local interest points obtained using a SIFT detector for human detection. These features were stored in a codebook for matching.

The second group is *dense* representation, where features are extracted densely over a detection window and concatenated into a high-dimensional descriptor. Various features such as Edgelet [5], Histogram of Oriented Gradients (HOG) [6], HOG with shape template [7], Covariance descriptors [8], Local Binary Pattern (LBP) and its variants [9, 10], Extended Histogram of Gradients [11, 12] and Histogram of Template [13] have been proposed over the recent years. A common trait among most of these features is that they are

usually represented by image intensities or gradients. This is not surprising as the contour of humans contains discriminative information that enables its differentiation from non-human objects.

LBP is a computationally-efficient texture feature that is popular in texture classification [14, 15, 16, 17] and face detection [18, 19]. LBP is illumination and contrast invariant as it only considers the sign of the difference between two pixels. Representing LBP in the form of histogram makes the descriptor resistant to translations within the neighbourhood of histogramming. There are works that propose variants of LBP for human detection. In [9], the authors identified the storage requirement of LBP features and the huge dissimilarity of semantically similar features as limitations of LBP for human detection. As such, they proposed 2 variants of LBP to better describe humans. In [10], the authors proposed using a concatenation of cell-structured LBP, similar to [6], and HOG to describe humans. Their feature showed a much better performance compared to [9]. Nguyen *et al.* [4] proposed using NRLBP for human description by mapping an LBP code and its complement to the minimum of the two.

However, for human detection, LBP has two issues. Firstly, LBP differentiates a bright human against a dark background and vice versa. This differentiation is undesirable in the context of human detection. In [4], NRLBP is proposed to map an LBP code and its complement to the minimum of the two to solve the problem. However, this presents another issue. In the *same block*, NRLBP maps a LBP code and its complementary representation to the same value. This produces a similar feature for some different local structures. Hence, NRLBP is unable to differentiate some local structures. Lastly, being illumination and contrast invariant, LBP and NRLBP do not differentiate between a weak contrast local region and a similar strong contrast one. Human contours tend to be situated in regions of strong contrast. Therefore, by totally discarding contrast information, the human contour may not be effectively discriminated by these features.

In this paper, we propose a novel edge-texture feature, Discriminative Robust LBP (DRLBP), in a dense representation to better discriminate between humans and non-humans. To alleviate the problems of LBP and NRLBP, the gradient weighted sum and absolute difference of a LBP code and its complement is considered. A linear SVM classifier is used for classification. The contribution of the paper is three-fold: 1) DRLBP differentiates most patterns which NRLBP misrepresents. Hence, we resolve the limitation of NRLBP whereby LBP codes and its complementary codes are mapped to the same code in the same block; 2) DRLBP, like NRLBP, does not differentiate the situations of a bright human against a dark background and vice versa; 3) DRLBP differentiates a weak contrast local region and a similar strong contrast one which enables it to represent human

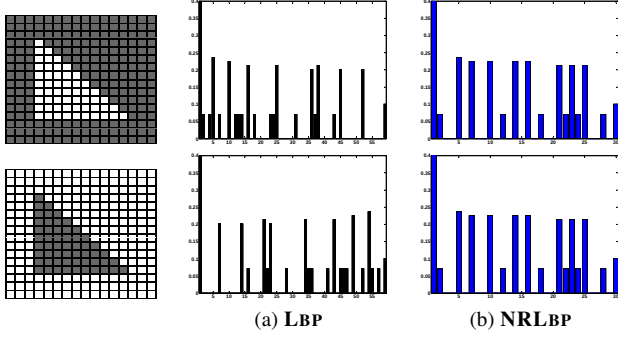


Fig. 1: Problem of Local Binary Patterns (LBP) and its solution by Non-Redundant LBP (NRLBP). For illustrative purposes, $\omega_{x,y} = 1$ and the vertical axis is truncated to make the bins clearer as there are uniform regions in the examples which results in high values in some bins.

contours more effectively compared to LBP and NRLBP.

2. DISCRIMINATIVE ROBUST LOCAL BINARY PATTERN

2.1. Limitations of LBP and NRLBP

The Local Binary Pattern (LBP) [16] code for a pixel at location (x, y) is computed as follows:

$$LBP_{x,y} = \sum_{b=0}^{B-1} s(p_b - p_c) 2^b, \quad (1)$$

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

where p_c is the value of the pixel at (x, y) , p_b is the value of the pixel in the b -th location on the circle of radius R around p_c and B is the total number of neighbouring pixels. For a $M \times N$ block, a LBP histogram of 2^B bins is computed for feature representation. There are some patterns that occur more frequently than others and the number of state transitions between 0 and 1 for these patterns are at most two [16]. Such patterns are termed as uniform patterns and the rest as non-uniform. By giving each unique uniform pattern a bin and collating all non-uniform patterns into a single bin, the number of bins for the histogram is reduced accordingly. For $B = 8$, the number of bins is reduced from 256 to 59. LBP is invariant to any monotonic changes to the image due to its fixed threshold of 0. Hence, it is illumination and contrast invariant.

An issue with LBP is that it differentiates a bright human from a dark background and vice-versa as the codes for the 2 situations are different. This differentiation makes the intra-class variation of the humans larger. In Fig. 1(a), the 2 situations in a block are illustrated for LBP. As it can be seen, the LBP features for the 2 situations are completely different.

A solution for the problem of LBP is proposed in [4] as Non-Redundant LBP. The authors propose mapping an LBP code and its complement to the minimum of the two. For instance, a LBP code of “1101 0101” and its complement, “0010 1010” will be treated as “0010 1010” in the mapping. Hence, the code “1101 0101” becomes redundant as it is never used in the histogram. NRLBP is robust to the reversal in intensity between the background and the humans than LBP. NRLBP can be computed as follows:

$$NRLBP_{x,y} = \min \{ LBP_{x,y}, 2^B - 1 - LBP_{x,y} \}, \quad (2)$$

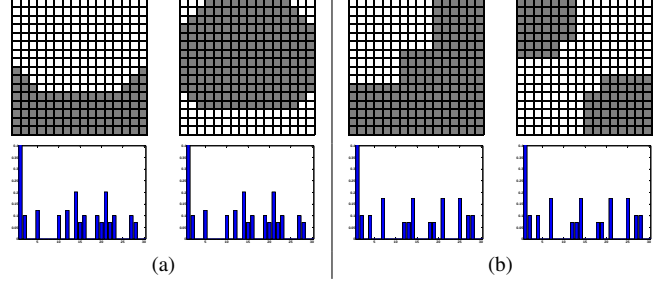


Fig. 2: Problem of Non-Redundant LBP (NRLBP). 4 local structures are shown in the first row. The second row shows the NRLBP feature for each structure. For each structure pair, the NRLBP feature is the same. For illustrative purposes, $\omega_{x,y} = 1$ and the vertical axis is truncated to make the bins clearer as there are uniform regions in the examples which results in high values in some bins.

Since the mapping reduces the number of codes by half, the number of bins for NRLBP histogram is 128 for $B = 8$. Using uniform codes, the number of bins is further reduced to 30. Fig. 1(b) illustrates how NRLBP mitigates the brightness reversal issue of human and background of LBP. It can be observed that for both situations, the NRLBP feature is similar.

In order to alleviate the intensity reversal problem of human and background, NRLBP maps LBP codes to the minimum of the code and its complement. However, this mapping function makes it difficult for NRLBP to differentiate some local structures that are dissimilar. It is possible that 2 different structures may have a similar feature representation. This is illustrated in Fig. 2 in the second row. This problem of NRLBP is caused by merging complement codes *in the same block*.

2.2. The Proposed Discriminative Robust Local Binary Pattern

For human detection, the contour of the human, which typically resides in high contrast regions between the human and the background, contains discriminatory information. LBP is illumination and contrast invariant. The histogramming of LBP codes only considers the frequencies of the codes i.e. the weight for each code in the block is 1. This form of histogram is unable to differentiate between similar regions of different contrast. Therefore, a weak contrast local region and a strong contrast one have similar feature representations.

To mitigate this problem, a weighting scheme is proposed. Given an image window, following [6], the square root of the pixels is taken. Then, the first order gradients are computed in the x - and y -directions. The gradient magnitude at each pixel is then computed and used to weigh its LBP code. The stronger the contrast at the pixel, the larger the weight assigned to the LTP code at that pixel. Consider a LBP histogram for a $M \times N$ image block. The value of the i^{th} bin of the weighted LBP histogram is as follows:

$$h_{lbp}(i) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \omega_{x,y} \delta(LBP_{x,y}, i), \quad (3)$$

$$\delta(m, n) = \begin{cases} 1, & m = n \\ 0, & \text{otherwise} \end{cases}$$

where $\omega_{x,y}$ is the gradient magnitude at the pixel location (x, y) .

It is not difficult to see that the NRLBP histogram can be computed from (3) as follows:

$$h_{nrlbp}(i) = h_{lbp}(i) + h_{lbp}(2^B - 1 - i), \quad 0 \leq i \leq 2^{B-1} - 1 \quad (4)$$

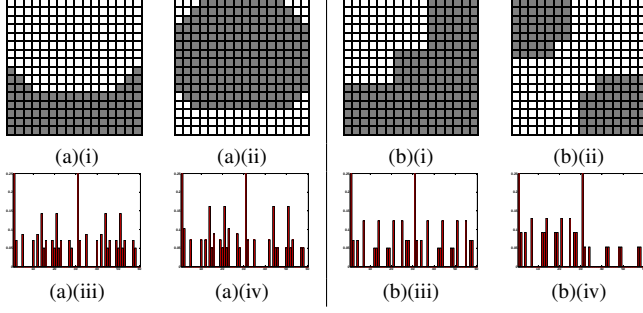


Fig. 3: DRLBP differentiates the local structure pairs in Fig. 2 mis-represented by NRLBP. For illustrative purposes, $\omega_{x,y} = 1$ and the vertical axis is truncated to make the bins clearer as there are uniform regions in the examples which results in high values in some bins.

where $h_{nrlbp}(i)$ is the i^{th} bin value of NRLBP. In order to resolve the issue of NRLBP whereby, in the same block, all LBP codes and their complements are mapped to the same bin, the following is proposed. Consider the absolute difference between the bins representing a LBP code and its complement to form Difference of LBP (DLBP) block histograms as follows:

$$h_{dlbp}(i) = |h_{lbp}(i) - h_{lbp}(2^B - 1 - i)|, \quad 0 \leq i \leq 2^{B-1} - 1 \quad (5)$$

where $h_{dlbp}(i)$ is the i^{th} bin value of DLBP. For blocks that contain structures that have both LBP codes and their complements, DLBP assigns small or almost zero values to the bins that the codes are being mapped to. By doing so, it differentiates these structures from those having no complement codes.

The 2 histogram features, NRLBP and DLBP, are concatenated to form *Discriminative Robust LBP* (DRLBP). The value of the i^{th} bin of DRLBP histogram is as follows:

$$\begin{aligned} h_{drlbp}(i) &= \begin{cases} h_{nrlbp}(i), & 0 \leq i \leq 2^{B-1} - 1 \\ h_{dlbp}(i - 2^{B-1}), & 2^{B-1} \leq i < 2^B \end{cases} \\ &= \begin{cases} h_{lbp}(i) + h_{lbp}(2^B - 1 - i), & 0 \leq i \leq 2^{B-1} - 1 \\ |h_{lbp}(i) - h_{lbp}(2^B - 1 - i)|, & 2^{B-1} \leq i < 2^B \end{cases} \end{aligned} \quad (6)$$

For $B = 8$, the number of bins is 256. Using uniform pattern representation, the number of bins is reduced to 60. Fig. 3 illustrates how DRLBP produce unique features for the structures shown earlier in Fig. 2. Hence, DRLBP represents the human contour more discriminatively than LBP and NRLBP. DRLBP also resolves the issue of differentiation of a bright object against a dark background and vice-versa as shown in Fig. 4.

3. EXPERIMENTAL STUDY

We perform experiments on two challenging data sets - INRIA [6] and Caltech Pedestrian Data Set [20]. Results are reported for both data sets using the per-image methodology suggested in [20] as the authors have shown it to be a better evaluation method. The per-image performance for dense representations of LBP and NRLBP on INRIA and Caltech, to the best of our knowledge, have not been published to date. Hence, experiments are performed for these features on the INRIA and Caltech data set. The feature parameters for LBP, NRLBP and DRLBP are set as follows. For both data sets, a block size of 16×16 pixels is used. A neighbourhood of 8 (B) pixels is considered using a circle of radius 1 (R). A 50% overlap

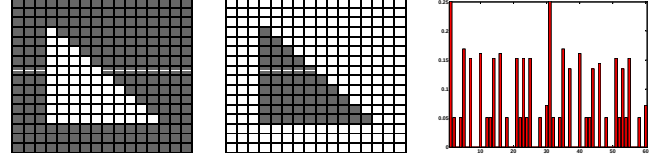


Fig. 4: Same DRLBPs are produced for patterns in Fig. 1. For illustrative purposes, $\omega_{x,y} = 1$ and the vertical axis is truncated to make the bins clearer as there are uniform regions in the examples which results in high values in some bins.

of blocks is used in the construction of the features. Square root of L1 norm normalization is used as our preliminary experiments show that this gives the best results. The overlapping block features for the image window is concatenated to form the overall window feature for training the linear SVM classifier.

The training set of INRIA is used to train the classifiers for INRIA and Caltech Data Sets. The training data set contains 2416 cropped positive training images and 1218 uncropped negative training images. The sliding image window size is 128×64 pixels. We randomly take 10 samples from each negative image to obtain a total of 12180 negative samples for training the linear SVM classifier. Bootstrapping is then performed on the negative images across multiple scales at a scale step of 1.05 to obtain hard negatives which are combined with the original training set to retrain the SVM classifier. This training procedure is exactly the same as the ones described in [6] and [20]. Note that the NRLBP feature representation in our work differs from the one in [4]. In our work, dense representation is adopted while sparse representation was used in [4] with probabilistic classification.

3.1. Results on INRIA Data Set

The INRIA test set consist of 288 images. The images are scanned using over multiple scales at a scale step of 1.05. The window stride is 8 pixels in the x and y directions. These parameters are the same as those used in [20] for test. The miss rate against false positives per image (FPPI) is plotted to compare between different detectors. The *log-average miss rate* [20] is used to summarize the detector performance.

The performance of DRLBP is compared with LBP, NRLBP and HOGLBP [10] (best performing LBP variant) in Fig. 5. The results of HOGLBP is given in [20]. This detector is trained and optimized by [10] and tested in [20]. DRLBP achieves a log-average miss rate of 36% which is significantly lower than all the methods being compared with. HOGLBP has a log-average miss rate of 39%. NRLBP performs worse than LBP as there is a significant loss of information due to the mapping of the LBP code and its complement to the same value. DRLBP consistently outperforms LBP and NRLBP. It even outperforms HOGLBP which is a hybrid feature representation. Fig. 7a shows some examples of DRLBP detections.

3.2. Results on Caltech Pedestrian Detection Benchmark Data Set

The Caltech Pedestrian Detection Benchmark testing data set contains 155 000 annotated pedestrian samples in 65 000 images and 56 000 negative images. The authors in [20] reported results whereby they used detectors trained on other data sets like INRIA for classification on their test set. Here, the results are also presented in a similar manner to [20] where detectors are trained using the INRIA data set and tested on the test set of Caltech. The scale step used is 1.05. The window stride is 8 pixels in the x and y directions. The

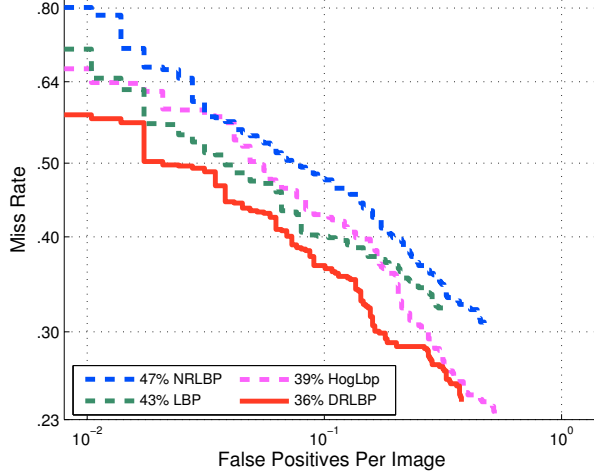


Fig. 5: Performance of DRLBP against LBP, NRLBP and HOGLBP [Best viewed in colour]. DRLBP outperforms all other methods.

settings used here are similar to those used in [20]. Similar to [20], in order to detect humans at smaller scales, the original images are upsampled. Only every 30th frame is evaluated so that comparisons can be kept consistent with those in [20].

The detectors we compare with our implemented detectors are the same as those in Section 3.1. The miss rate versus false positives per image [20] is plotted and log-average miss rate is used as a common reference value for summarizing performance. Fig. 6 plots the performance on 50-pixel or taller, unoccluded or partially occluded pedestrians (reasonable evaluation setting). DRLBP performs the best at a log-average miss rate of 62%. HOGLBP has a log-average miss rate of 68% while LBP and NRLBP have log-average miss rate of 79% and 85% respectively. DRLBP consistently outperforms LBP and NRLBP. At values of 10⁻² FPPI and higher, DRLBP outperforms HOGLBP consistently. Fig. 7b shows some examples of DRLBP detections.

3.3. Relation to prior work

The work in this paper focuses on a novel edge-texture feature, DRLBP, which fuses gradient and texture information into a single feature. Although Nguyen *et al.* [4] identified a limitation of LBP whereby a bright human against a dark background is differentiated from a dark human against a bright background, their proposed feature, NRLBP, has limitations. NRLBP maps LBP codes and its complements in the same block to the same code. This causes some structures to be misrepresented by NRLBP. Furthermore, NRLBP, like LBP, discards contrast information. As a result, similar regions with different contrast have similar feature representation. HOGLBP [10] is the most closely-related and best performing LBP variant feature to date. However, HOGLBP is made up of concatenated edge (HOG) and texture (LBP) features. Since LBP was not modified, HOGLBP suffers from the same problems identified for LBP. Our work capitalizes on the limitations of LBP and NRLBP which was not considered by any of the previous work [4, 9, 10]. Our proposed feature, DRLBP, considers both the gradient weighted sum and absolute difference of the bins of the LBP codes with their respective complement codes.

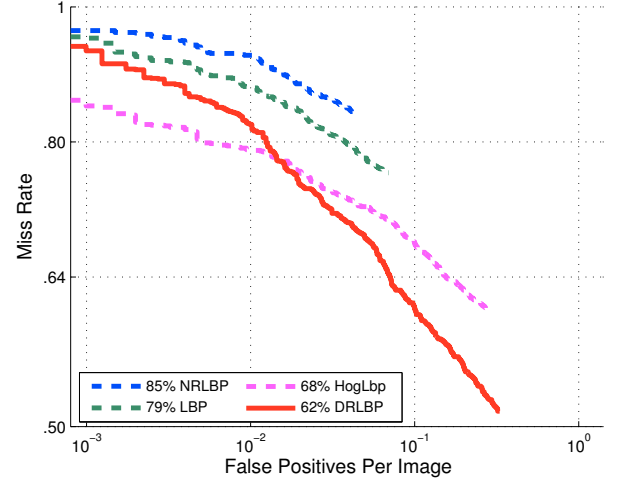


Fig. 6: Evaluation results under reasonable condition on test set of Caltech Pedestrian Data Set [Best viewed in colour]. DRLBP outperforms all other methods.

4. CONCLUSION

This paper presents a novel edge-texture feature, Discriminative Robust Local Binary Pattern (DRLBP), for human detection to alleviate the limitations of Local Binary Pattern (LBP) and Non-Redundant LBP (NRLBP). LBP differentiates a bright human against a dark background and vice-versa. For human detection, this differentiation makes the intra-class variation of humans larger. NRLBP mitigates the problem of LBP by choosing the minimum of a LBP code and its complement. However, NRLBP maps LBP codes and its complements in the same block to the same code. This causes some structures to be misrepresented by NRLBP. Furthermore, LBP and NRLBP discard contrast information in their representation. As a result, similar regions with different contrast have similar feature representation. For human detection, this is not desired as the human contour contains the most relevant information. By ignoring the contrast information, the contour is not effectively discriminated by the features. The new feature, DRLBP, considers both the gradient weighted sum and absolute difference of the bins of the LBP codes with their respective complement codes. In this way, DRLBP alleviates the problems of LBP and NRLBP for human detection. Results demonstrate that DRLBP always outperforms LBP and NRLBP consistently and performs better than HOGLBP.

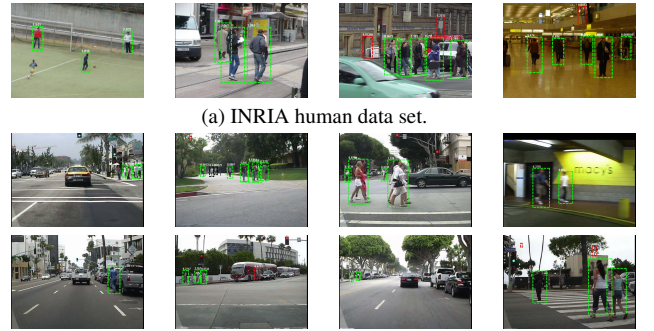


Fig. 7: Detection results of DRLBP on various data sets used in the experiments [Best viewed in colour]. True detections are indicated in green while false positives are in red.

5. REFERENCES

- [1] S. Ioffe and D.A. Forsyth, "Probabilistic methods for finding people," *Int. J. Comput. Vis.*, vol. 43, pp. 45–68, 2001.
- [2] R. Ronfard, C. Schmid, and B. Triggs, "Learning to parse pictures of people," in *Proc. Eur. Conf. Comput. Vis.*, vol. 2353, pp. 700–714, 2002.
- [3] A. Opelt, A. Pinz, and A. Zisserman, "Incremental learning of object detectors using a visual shape alphabet," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, June 2006, vol. 1, pp. 3 – 10.
- [4] N. Duc Thanh, Z. Zhimin, P. Ogunbona, and L. Wanqing, "Object detection using non-redundant local binary patterns," in *Proc. IEEE Int. Conf. Image. Process.*, Sept. 2010, pp. 4609 –4612.
- [5] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. I: 90–97.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. I: 886–893.
- [7] L. Zhe and L.S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 604 –618, Apr. 2010.
- [8] O. Tuzel, F.M. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
- [9] M. Yadong, Y. Shuicheng, L. Yi, T. Huang, and Z. Bingfeng, "Discriminative local binary patterns for human detection in personal album," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, June 2008, pp. 1 –8.
- [10] T.X. Wang, X. Han and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 32–39.
- [11] A. Satpathy, X.D. Jiang, and H-L. Eng, "Extended histogram of gradients feature for human detection," in *Proc. IEEE Int. Conf. Image. Process.*, Sept. 2010, pp. 3473 –3476.
- [12] A. Satpathy, X.D. Jiang, and H-L. Eng, "Extended histogram of gradients with asymmetric principal component and discriminant analyses for human detection," in *Proc. IEEE Canad. Conf. Comput. Robot. Vis.*, May 2011, pp. 64 –71.
- [13] S. Tang and S. Goto, "Histogram of template for human detection," in *Proc. IEEE Int. Conf. Acous. Speech. Sig. Process.*, Mar. 2010, pp. 2186 –2189.
- [14] G. Zhenhua, Z. Lei, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657 – 1663, June 2010.
- [15] S. Liao, M.W.K. Law, and A.C.S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1107 – 1118, May 2009.
- [16] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971 –987, July 2002.
- [17] Z. Guoying and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915 – 928, June 2007.
- [18] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037 – 2041, Dec. 2006.
- [19] T. Xiaoyang and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635 –1650, June 2010.
- [20] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743 –761, Apr. 2012.