LETTER

A Novel Pedestrian Detector on Low-Resolution Images: Gradient LBP Using Patterns of Oriented Edges

Ahmed BOUDISSA†, **Joo Kooi TAN**†, ***Nonmembers***, **Hyoungseop KIM**†, ***Member***, **Takashi SHINOMIYA**††, ***Nonmember***, ***and*** **Seiji ISHIKAWA**†, ***Member***

SUMMARY This paper introduces a simple algorithm for pedestrian detection on low resolution images. The main objective is to create a successful means for real-time pedestrian detection. While the framework of the system consists of edge orientations combined with the local binary patterns (LBP) feature extractor, a novel way of selecting the threshold is introduced. Using the mean-variance of the background examples this threshold improves significantly the detection rate as well as the processing time. Furthermore, it makes the system robust to uniformly cluttered backgrounds, noise and light variations. The test data is the INRIA pedestrian dataset and for the classification, a support vector machine with a radial basis function (RBF) kernel is used. The system performs at state-of-the-art detection rates while being intuitive as well as very fast which leaves sufficient processing time for further operations such as tracking and danger estimation.

key words: *local binary patterns, pedestrian detection, object recognition, support vector machine*

1. Introduction

In the interest of making roads safer and allowing driver assistance systems to avert accidents due to lack of attention, many researchers acknowledged a crucial step in this process namely pedestrian /human detection [1], [6], [9], [15], [16] and [19]. Although, in the past, many algorithms have tackled this problem, and some of them performed very efficiently- insofar as the detection rate is concerned- there are only a few of these methods that are applicable in practice, mainly because of their computational complexity. Against this background, we hereby introduce a novel method for pedestrian detection. Inspired by the patterns of oriented edge magnitudes introduced in [5], that is a spatial multi-resolution descriptor that captures rich information about the original image. It makes use of the gradient magnitude orientations, before applying the local binary patterns (LBP) operator. This feature detector has proven its efficiency in face detection schemes by outperforming the major approaches such as Local Gabor Binary Patterns [7] and Histograms of Gabor Phase Patterns [10] on both detection rate and processing speed. Our framework originality resides within the extraction of single scale gradient local binary patterns computed over different orientations. This descriptor is tested on low resolution images (30×60 pixels) of the INRIA pedestrian dataset [11] available online. As will be further illustrated in this paper, it was found that the system performed best when not using the gradient histograms, and therefore a new scheme to select the Local binary patterns (LBP) threshold τ using the variance of the negative samples is introduced.

Manuscript received February 1, 2013.
Manuscript revised July 1, 2013.

†The authors are with Kyushu Institute of Technology, Kitakyushu-shi, 804–8550 Japan.
††The author is with Japan University of Economics, Tokyo, 150–0031 Japan.

DOI: 10.1587/transinf.E96.D.2882

2. Related Work

Lately, there has been a marked interest in local descriptors and their application in object recognition, specifically in pedestrian detection. Mikolajczyk and Schmid [2] provide an extended survey on the different local descriptors and their evaluation. As for pedestrian detection, Enzweiler and Gavrila [3] as well as Gandhi [4] provide in-depth surveys and evaluation of the state-of-the-art pedestrian detection frameworks.

Feature based approaches using local filtering on image different locations (image cells or single pixels) are popular in pedestrian detection such as Papageorgiou and Poggio's non-adaptive Haar-wavelet [8]. This approach is solely based on a dense feature dictionary representation of the intensity difference in a multi-scale and orientation fashion and the use of the integral images [12] made this feature set both popular [13], [20], [23] and simple to evaluate.

However, the multiple redundancy of these features required a feature selection mechanism; either manually using a prior knowledge about the human body geometry [8], [13], [14], or using the boosting technique [17] and its variants (Adaboost [20]) which select automatically the most discriminative features by generating a strong classifier out of a set of weak classifiers.

Another class of local feature extractors that has been used in pedestrian detection is the codebook feature patches. As in [18], [21] and [26], generated from the training data, these features are extracted around the points of interest in the image, and coded along with their spatial relationship into a feature vector.

Recently, and closer to our own framework, there has been an interest toward local edge descriptors, such as local gradient histograms extracted from normalized gradient images over "blocks" and their combination with the edge orientation. In fact, "HOG-like" features received a special interest in various works [1], [22] and [24].

Our approach is feature- based, inspired by [5], and based on Ojala et al.'s local binary pattern feature

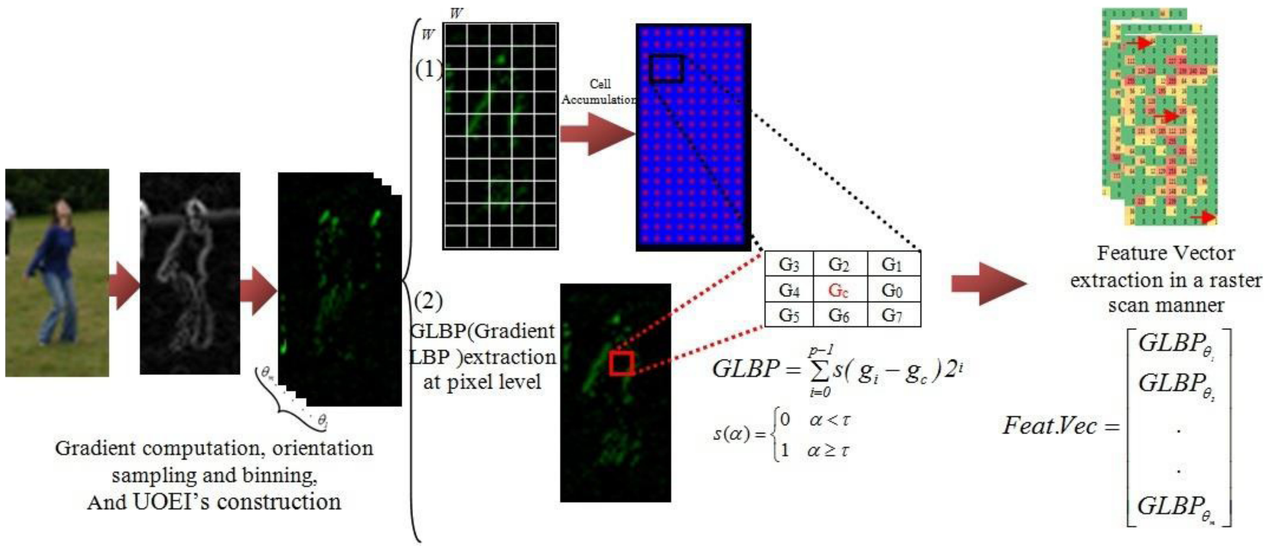


Fig. 1 Overview of the approach: (1) Extracting LBP using cell accumulation (2) pixel wise Lbp-extraction. Then the feature vector extraction is a raster scan manner which is fed to the non-linear SVM.

extractor [25]. The LBPs have shown to be efficient in texture classification [29] and face detection and recognition [27]. It also inspired many pedestrian detection studies as it is intuitive and easy-to-implement [28], [30]. Combined with the histograms of oriented gradients (HOG) [1] in [28] and an occlusion handling scheme, it has shown notable results on the INRIA pedestrian dataset.

Other LBP variants are the center-symmetric local binary patterns (CS-LBP) and pyramid center-symmetric local binary/ternary patterns (CS-LBP/LTP) [30]. They capture the gradient information and some texture information densely over multiple scales and yield good results on the same dataset.

What can be argued is that, despite the high performance obtained in these approaches, the augmented complexity of the feature detectors (even if optimized properly) might result in a real-time detector, but will still limit any further processing perspectives, which is in the heart of our scope. Throughout the literature, we see clearly that the feature extraction step is an important one, on which the quality of the detector depends directly. Features must be discriminative and robust. In [31], it is stated that the LBP (classical) [29] is not well-suited for pedestrian detection, and that is why we are presenting a modified version. Also, in [30], it is argued that HOG [1]-like features (based on histograms of oriented edges) are likely to perform badly within real situations of pedestrian detection owing to the sensitivity of the gradient operator to noisy and cluttered backgrounds. In this paper, we will demonstrate that the use of a single scale dense feature extraction and a proper choice of the LBP threshold resolve partially the problems of edge based methods, leading to the construction of a compact descriptor that inherits various good properties from existing features with a low computational cost and a state-of-the-art accuracy on the INRIA dataset.

3. Overview

The main objective of this work is to take over the following challenges:
- Use the LBP computational efficiency to design a new feature detector, simpler and easier to compute than the existing methods.
- Detect pedestrians on low resolution images (Height less than 40 pixels). In order to underline the importance of an early detection in a car vision system, giving more time for preventive action.

The key idea here is to make use of the gradient to extract the shape information, adding to this the orientation information of the edges giving a rich description of the pedestrian shape. Using LBP, self-similarities of the edges are encoded within a feature vector. As shown in Fig. 1, we first compute the gradient images using a simple gradient operator: $[-1\ \ 0\ \ 1]$ on x and y directions. Every pixel is then replaced by the value of the gradient at this location. The gradient orientations are sampled into m bins. We use unsigned gradient orientation (0°-180°). After that, we construct m images, called uni-orientation edge images (UOEI) from the original gradient one, by assigning every pixel p with a gradient orientation θ_m to the image m.

For the next step we adopt two approaches:

1- Assigning every pixel the sum of its $w * w$ cell (Cell Accumulation), and then using the LBP operator over cell blocks Bl.
2- Applying the LBP operator directly to the different orientation images.

For the first approach, we varied accumulation cell size $w = \{3, 5, 7, 9\}$, bin number $m = \{3, 4, 5, 6, 7, 8, 9\}$ and block size Bl. The block size is set to $2w$ for no cell overlap and to

$2w-2$ for an overlap of the last row with the first row of the adjacent cell (See Fig. 1). For the second approach, a dense LBP characterization was used and two different block geometries were tested: rectangular and circular.

4. The Local Binary Patterns Threshold

The local binary pattern (LBP Fig. 1) features are a self-similarity measure that was introduced for the first time in 1996 by Ojala et al. [29]. Owing its simplicity and efficiency, it became popular and found new horizons in many applications. A detailed LBP-related bibliography can be found online [33].

Our choice for the LBP features is not a trivial one, it has proven to be highly discriminative and its key advantages, namely its invariance to monotonic gray level changes and computational efficiency, make it suitable for pedestrian detection.

There has been few works applying the LBP features to object detection frameworks, mainly face recognition [27], [29]. In the original LBP, the goal is to describe every texture and face in a discriminative way, so the LBP threshold was set to 0. We can see in Fig. 2. The LBP of the patch from the background (i.e. grass), if treated with the classical local binary patterns, produces random patterns. These patterns are hard to distinguish from the patterns making up the human shape, which we are trying to extract. In Wolf et al. [32], it is stated that for a good stability of the LBP operator in uniform regions, the threshold should be chosen to be close to zero ($\tau = 0.01$).

In our application, namely pedestrian detection, the goal is to create a robust model for pedestrians. In order to achieve that, we introduce a new way to select the LBP threshold. We base our choice of the LBP threshold on the following conjecture:

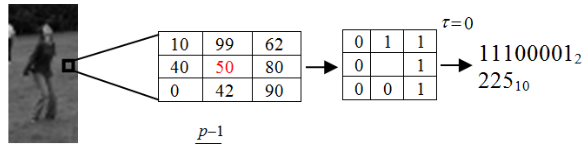
*"In a car vision application, the background on which pedestrians appear, i.e., negative samples, is **mostly** uniformly cluttered regions such as road asphalt, trees, sky, grass and mud.* (Fig. 3)"

Based on this conjecture, the choice of the threshold is taken to be equal to the average variance of the gradient value of the negative samples in the database defined by Eqs. (1) & (2).

Since our approach is solely based on pedestrian shape extraction, the reasoning behind choosing the variance is to set a value for this threshold to eliminate what can be informally referred to as "weak edges" which constitute the background, while preserving the "strong edges" describing the pedestrian shape. Furthermore, ignoring the background details results in more zero's within the feature vector, which adds to the computational efficiency to the classifier.

5. Training and Results Discussion

A non-linear support vector machine is trained and used for the classification [34]. We use a radial basis function kernel (RBF). As for our approach in this paper, the training is done offline. Also, with the simplicity that characterizes the feature we use in this study, we can say that a non-linear SVM can be afforded processing time wise. The training and test data- the INRIA dataset [11]-consists of a 1,208 pedestrian images training set and a test set of 288 images with 589 human samples and 453 human free images. For the training data, the original image size is 96×160 pixels with a margin of 16 pixels around each side. In an attempt in our work to seek the challenge of small pedestrians, the images have been cropped and scaled to 30×60 pixels.



$$LBP = \sum_{i=0}^{p-1} s(g_i - g_c)2^i \qquad s(\alpha) = \begin{cases} 0 & \alpha < \tau \\ 1 & \alpha \geq \tau \end{cases} \tag{1}$$

Fig. 2 Local binary patterns (LBP) [29].

$$\tau = Var = \sum_{i=1}^{N} \frac{|Var_i|}{N} \tag{1}$$

$$Var_i^2 = \sum_{k=0}^{W-1} \sum_{l=0}^{H-1} \frac{Gr_i[k][l] - \overline{Gr_i}}{W \times H} \tag{2}$$

N is the number of negative samples, $Gr_i[k][l]$ is the gradient value at location (k,l). Var_i is the gradient variance of image i.

Fig. 3 Random patches sample taken from INRIA negative samples.

In [5], the patterns of oriented edge magnitude applied to face recognition show that the best performance is achieved with the number of bins $m = 3$, which is understandable since, in their application, the main characteristic of a feature vector is discriminability. On the other hand, in human detection, we need a robust model immune to noise with a rich shape description. All the results that are presented in the following section employ unsigned gradient orientation with $m = 9$.

Many settings have been tested, and in what follows we will present the most relevant ones:

5.1 Window Size $w = \{3, 5, 7, 9\}$:

From the graph on (Fig. 5), we see that the larger the accumulation window, the weaker is the system's performance, which can be explained by the fact that details tend to be smoothed away, and information dimmed and lost since the images we are working on are low resolution images. This provides us with the main advantage; which is applying the

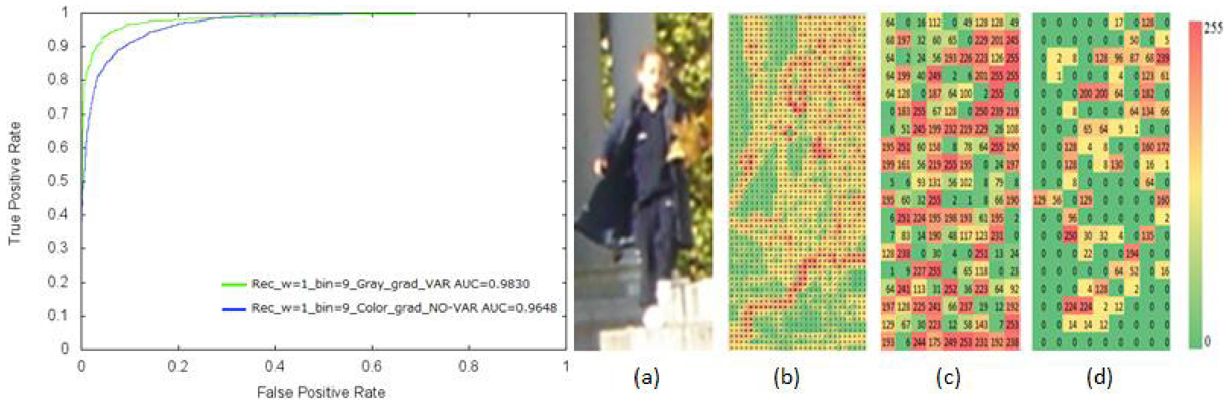


Fig. 4 (Left) Performance curves for classical LBP threshold versus the proposed value. (Right) Qualitative comparison of the effect of the new threshold choice versus the old classical LBP zero threshold. (a) Original Image, (b) gradient energy image, (c) LBP-image with $\tau = 0$, (d) LBP image with $\tau = \text{Var}$.
REC = Rectangular LBP Neighborhood w = Accumulation window size: w = 1 means no accumulation Bin = bin number m Gray/Color_grad= color gradient as compared to gray scale gradient. Var = threshold τ equal to variance NO-VAR = LBP threshold τ = 0

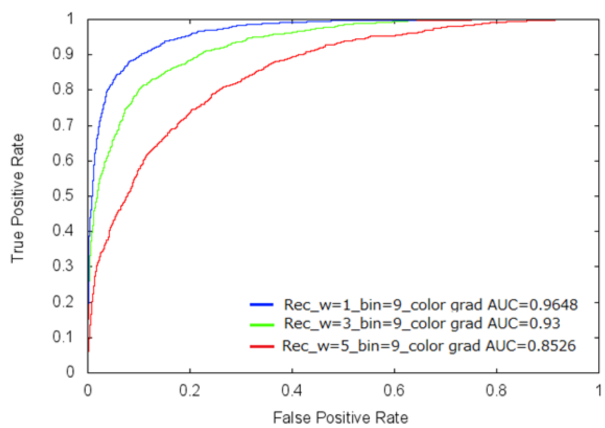


Fig. 5 Different performance curves when varying the accumulation window size *w*- *REC = Rectangular LBP Neighborhood w = Accumulation window size: w = 1 means no accumulation Bin = bin number m Gray/Color_grad = color gradient as compared to gray scale gradient.*

LBP operator directly. Also, by eliminating the need to compute both the integral image and the spatial accumulation, we achieve a shorter path to our algorithm, which improves the computation time significantly.

5.2 Classical Threshold ($\tau = 0$) vs. New Threshold ($\tau = Var$)

This choice represents the main originality that this study brings to the field. The previously depicted results were for a classic threshold $\tau = 0$ (*NO-Var*) [25]. In Fig. 4 (left), we can see that the use of the variance of the negative components improves the discriminative power of this feature extractor and practically improves the performance, as it eliminates more than 5% of the false positives.

In Fig. 4 (right), we can see the effect of using our proposed estimation of the threshold. Figure 4-c- shows the LBP image extracted using the classical threshold. We can see that the background clutter produces noise patterns which remain indiscernible from the useful shape information we are to extract. Figure 4-d shows the LBP image resulting from the application of the variance threshold. We can see that the human shape is perceptible; while the background clutter has been discarded (green areas are zeros). The use of the proposed threshold targets uniformly cluttered areas as explained in Sect. 4. This result confirms the conjecture stated earlier.

In order to improve efficiency, and to get the best performance possible from this approach, we also tried the circular LBP block geometry, using the bilinear interpolation techniques when the desired pixel value does not lie on the center of a pixel. We noticed a slight improvement of the performance. This is due to the homogeneity of the neighborhoods when all the pixels of the LBP block are at the same distance from the center pixel.

Also, to minimize the false positives, we used what is referred to as “retraining.” This technique consists of training a model with the available negative/positive samples, after which we took 10,000 random 30×60 pixel patches from the negative images, using them as test data, and retraining the original model by adding the false positives that have been generated. By adding those “hard examples” to the training model, the recall improved by 6%. In Fig. 6 we depict both the effect of applying the circular neighborhoods as well as the retraining. For the highest curve, with a circular neighborhood, and with orientations sampled to 9 directions, using gray scale gradient, combined with the threshold $\tau = Var$ and with retraining, we obtained an accuracy of 98.44% at 10^{-2} false positives per window (FPPW).

In the next graph, Fig. 7, we depict a comparison of our approach and two of the most similar frameworks (dense feature extraction). Our approach clearly outperforms the classical histograms of oriented edge gradients. Also, it

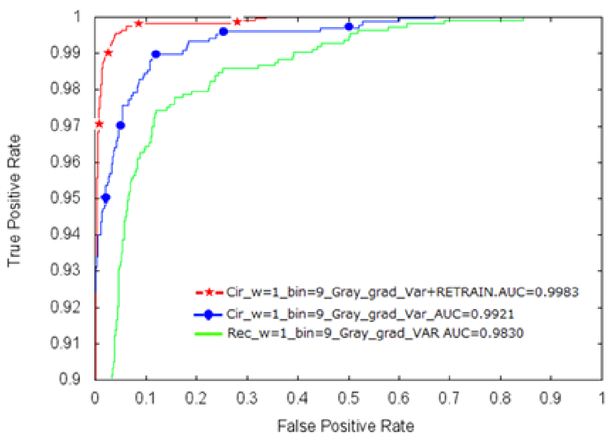


Fig. 6 The combined effect of retraining and circular LBP neighbourhood. *REC = Rectangular LBP Neighborhood w = Accumulation window size: w = 1 means no accumulation Bin = bin number m Gray/Color_grad = color gradient as compared to gray scale gradient. Var = threshold τ equal to variance NO-VAR = LBP threshold τ = 0.*

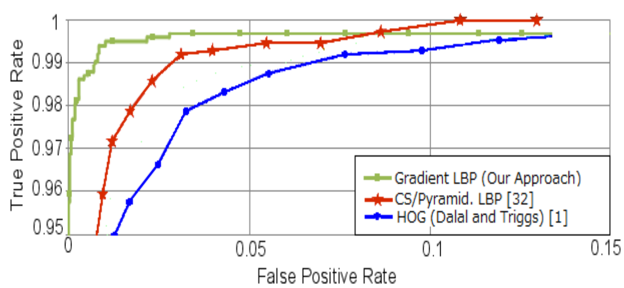


Fig. 7 Performance comparison between the main methods: Our approach, HOG [1] and CS/Pyramid LBP [31].

Fig. 8 Examples of correct detections by our method which have been missed by HOG [1] and Cs-Pyramid LBP [31].

Table 1 Algorithm flow length comparison between [31], [1] and the proposed approach.

	Proposed approach	Cs-LBP [32]	HOG[1]
Gradient Computation	✓	✓	✓
Classifier	✓	✓	✓
Local binary patterns	✓	✓	
Normalizations		✓	✓
Histograms computation		✓	✓
Integral Image			✓

outperforms a more recent pedestrian detection framework tested on the INRIA pedestrian dataset, which is the center symmetrical and pyramidal LBP [30].

By analyzing those results, we confirm what was said in Sect. 1, that HOG-like features fail within real situations of pedestrian detection owing to the sensitivity of the gradient operator to noisy and cluttered backgrounds.

We find that when using low resolution images, the shape of the pedestrians becomes harder to distinguish from the background which causes many methods to fail (Fig. 8.)

For the histograms of oriented gradient [1], [11], the feature vector is several tens of thousands of dimension. When dealing with small pedestrians, the discriminative details get lost in the high dimensionality of the feature vector. Also, for Cs-Lbp [30] the histogram nature of the feature vector makes the features being taken insensitive to small pedestrians. For both methods, the edges produced by the background are included in the feature vector, which, after the training, makes the distinction between useful information and noise difficult.

In our method, the feature vector is taken in a raster scan manner, which keeps the spatial relationship of the features. Also, it is about 1500 dimensional vector. The new threshold eliminates most of the background noisy patterns, which leads to a robust human model. Even with a limited amount of information, i.e. small pedestrians, using our method, pedestrians like in Fig. 8 could be detected efficiently.

Putting aside the classification comparison, the strongest point of this approach is the simplicity. As shown in Table 1 we can see that the flow of our algorithm is much shorter than the other methods. This is the main motivation behind this work, which is to reduce the complexity of pedestrian detectors, in order to include them in more complex frameworks.

6. Conclusion and Future Work

In this paper, we introduced a new feature descriptor to pedestrian detection. We found that replacing the classical zero LBP threshold by the variance of the negative samples quiets down locations which may initially have had a strong edge response, but which resemble their neighbors. It also has been shown that the drawback of the usual main edge based methods -the noisy response to cluttered backgrounds- can be overcome by using this threshold. State-of-the-art accuracy has been achieved, but above all, the simplicity of this feature opens up a new avenue for car-vision directed systems, since the most important part of any car vision system is the early detection and then an evaluation of the danger. We plan to implement this detector on a per-image level; a sliding window approach might be appropriate, but a more suitable approach would be considering the scene geometrical context.

Acknowledgments

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Grant Number 25350477.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," CVPR 2005, vol.1, pp.886–893, 2005.
[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.10, pp.1615–1630, 2005.
[3] M. Enzweiler and M. Gavrila, "Monocular pedestrian detection: Survey and Experiments," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.12, pp.2179–2195, Dec. 2009.
[4] T. Gandhi and M.M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," IEEE Trans. Intelligent Transportation Systems, vol.8, no.3, pp.413–430, Sept. 2007.
[5] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," ECCV 2010, pp.313–326, 2010.
[6] P. Doll´ar, B. Babenko, S. Belongie, P. Perona, and Z. Tu, "Multiple component learning for object detection," ECCV, 2008.
[7] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," ICCV, vol.1, pp.786–791, 2005.
[8] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.4, pp.349–361, April 2001.
[9] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "A mobile vision system for robust multi-person tracking," CVPR, 2008.
[10] B. Zhang, S.S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition," IEEE Trans. Image Process., vol.16, pp.57–68, 2007.
[11] N. Dalal, Finding People in Images and Videos, PhD thesis, Institut Nat'l Polytechnique de Grenoble, 2006.
[12] F. Crow, Summed-area tables for texture mapping, SIGGRAPH, 84, 207–212, 1984.
[13] H. Shimizu and T. Poggio, "Direction estimation of pedestrian from multiple still images," Proc. IEEE Intelligent Vehicles Symp., pp.596–600, 2004.
[14] C. Papageorgiou and T. Poggio, "A trainable system for object detection," Int. J. Comput. Vis., vol.38, pp.15–33, 2000.
[15] E. Seemann, M. Fritz, and B. Schiele, "Towards robust pedestrian detection in crowded image sequences," CVPR, 2007.
[16] D. Tran and D. Forsyth, "Configuration estimates improve pedestrian finding," NIPS, vol.20, pp.1529–1536, 2008.
[17] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Proc. European Conf. Computational Learning Theory, pp.23–37, 1995.
[18] B. Leibe, N. Cornelis, K. Cornelis, and L.V. Gool, "Dynamic 3D scene analysis from a moving vehicle," Proc. IEEE Comput. Vision Pattern Recognit., 2007.
[19] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: single-frame classification and system level performance," Intelligent Vehicles Symposium, 2004.
[20] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," Int. J. Comput. Vis., vol.63, no.2, pp.153–161, 2005.
[21] E. Seemann, M. Fritz, and B. Schiele, "Towards robust pedestrian detection in crowded image sequences," Proc. IEEE Comput. Vis. Pattern Recognit., 2007.
[22] V.D. Shet, J. Neumann, V. Ramesh, and L.S. Davis, "Bilattice-based logical reasoning for human detection," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2007.
[23] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," Proc. Int'l Conf. Image Processing, pp.900–903, 2002.
[24] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp.1491–1498, 2006.
[25] 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, 2002.
[26] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.11, pp.1475–1490, Nov. 2004.
[27] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," Proc. ECCV'04, pp.469–481, 2005.
[28] X.Y. Wang, X. Han, and S.C. Yan, An HOG-LBP human detector with partial occlusion handling, In ICCV, 2009.
[29] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognit., vol.29, pp.51–59, 1996.
[30] Y. Zheng, C. Shen, R. Hartley, and X. Huang, "Pyramid center-symmetric local binary/trinary patterns for effective pedestrian detection," Asian Conference on Computer Vision (ACCV), pp.281–292, 2011.
[31] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," CVPR, 2008.
[32] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," Real-Life Images Workshop at ECCV, 2008.
[33] http://www.cse.oulu.fi/MVG/LBP_Bibliography.
[34] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," Science 2.3, pp.1–39, 2001.