社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2011-247,HIP2011-75(2012-3)

Fast pedestrian detection using LBP-based patterns of oriented edges

Ahmed Boudissa, Joo Kooi Tan, Hyoungseop Kim, Seiji Ishikawa
Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Kitakyushu, Japan

Abstract— **This paper introduces a simple yet efficient algorithm for pedestrian detection on low resolution images. The main objective is to create a successful means of achieving a real-time pedestrian detection application. While the framework of the system consists of edge orientations combined with the LBP feature extractor, a novel way of selecting the threshold is introduced. With the objective being an efficient car vision algorithm, it is assumed that the negative samples in this context are mainly uniformly textured objects such as trees, roads, and buildings. 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 an RBF kernel is used. The kernel parameters are chosen to maximize the AUC of the receiver operating curve (ROC) . The system performs at a 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.**

I. Introduction

In the focus of making roads safer, and sustaining driver assistance systems to avert lack-of-attention accidents, many researches have seen the light concerning a crucial step of this process which is pedestrian /human detection [1, 6, 9, 15, 16, 19].

Although many algorithms tackled this problem for a moment, and some of them performed very efficiently-detection rate wise- there is only a few of these methods that are applicable practically, mainly because of their computational complexity. Even with Moore's law promising computational abilities in the future, there is an urging need for a practical real time pedestrian detection framework that needs to perform well on one side, and that leaves enough time for further processing beyond detection such as tracking and danger estimation which can themselves be time consuming.

For this perspective, we hereby introduce a novel method to pedestrian detection. Inspired by the patterns of oriented edge magnitudes introduced in [5], it is a spatial multi-resolution descriptor that captures rich information about the original image, making use of the gradient magnitudes orientations, afore applying the LBP operator. This feature detector has proven its efficiency in face detection schemes 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 (30x60 pixels) of the INRIA pedestrian dataset [11] available online. Also, and as it will be shown further, we found that the system performed best when not using the gradient histograms, and we introduce a new scheme to select the LBP threshold τ using the variance of the negative samples.

In section II of this paper, we will cover briefly the related works. In section III, our approach and the theoretical background that lies behind it will be detailed. In section IV, experiments are shown, and the training/testing data along with the evaluation process are explained. We wrap up this work with conclusions; remarks and future work as long as it is considered as a first step toward a practically implementable real-time car vision application.

II. Related work

Lately, there has been an extensive interest in local descriptors and their application in object recognition, and specifically in pedestrian detection. Mikolajczyk and Schmid [2] provide an extended survey on the different local descriptors and their evaluation. As for pedestrian detection, M. Enzweiler and M. Gavrila [3] and T. Gandhi [4] provide in-deep surveys and evaluation of the state-of-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

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.

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][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.

More recently, and closer to our 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 different works [1][22][24]. Whereas the HOG-like features are computed on a dense manner, SIFT [30] consists of an interest point detector that leads to a sparse representation.

Our approach is feature based, Inspired by [5], and based on Ojala et al.'s Local binary pattern feature 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 researches being intuitive, and easy-to-implement [28][32]. Combined with HOG[1] in [28] and an occlusion handling scheme, it has shown impressive results on the INRIA pedestrian dataset.

Another LBP variant is the center-symmetric local binary patterns (CS-LBP) and pyramid center-symmetric local binary/ternary patterns (CS-LBP/LTP) [32]. It captures the gradient information and some texture information densely over multiple scales and yielded good results on the same dataset.

What can be argued on these approaches is, despite their high performance, the augmented complexity of the feature detectors, which, even if optimized properly, can result in a real-time detector, but will still limit any further processing perspectives, which is in the heart of our scope.

Through 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 [33], it is stated that the LBP (classical)[29] is not well suited for pedestrian detection, and this is why we present here this modified version.

Also, in [32], it is argued that HOG-like features (edge based) are likely to perform badly within real situations of pedestrian detection due 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 resolves 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.

III. Overview

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.4**, 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- Following [5], 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 one, we varied accumulation cell size *w*= {3, 5, 7, 9}, bin number *m*= {3, 4, 5, 6, 7, 8, 9} and the block size *Bl* which 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. 3**).

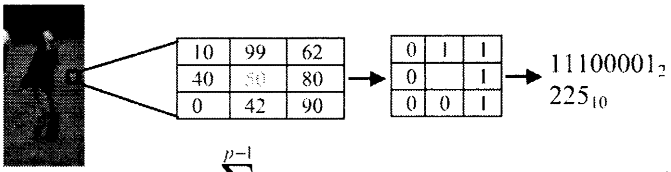
For the second approach, a dense LBP characterization was used and two different block geometries have been tested (rectangular and circular).

IV. The LBP operator

The Local binary pattern features are a self-similarity measure that has been introduced for the first time in 1996 by Ojala et al [29]. Due to its simplicity and efficiency, it became popular and found horizons in many applications. A detailed LBP-related bibliography can be found online [35].

Our choice for the LBP features is not trivial, as a matter of fact, 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.

Figure 1 shows the original LBP operator which assigns a label to every pixel of an image by thresholding the 3x3 square neighborhood of each pixel with the center pixel value (τ =0) and considering the result as a binary number.



10	99	62
40	50	80
0	42	90

0	1	1
0		1
0	0	1

$$LBP = \sum_{i=0}^{p-1} s(g_i - g_c)2^i \qquad (1)$$

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

Fig. 1. The classical LBP described in the original paper [29].

In *Wolf et al.* [34], 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 (τ =0.01).

To justify our choice of the LBP threshold, we assume 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* (See **Fig. 2**)*."*

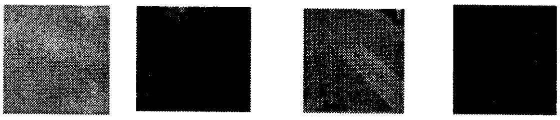


Fig. 2. Random patches sample taken from INRIA negative samples.

Based on this conjecture, the choice of the threshold τ is taken to be equal to the average variance of the gradient value of the negative samples in the database defined by

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

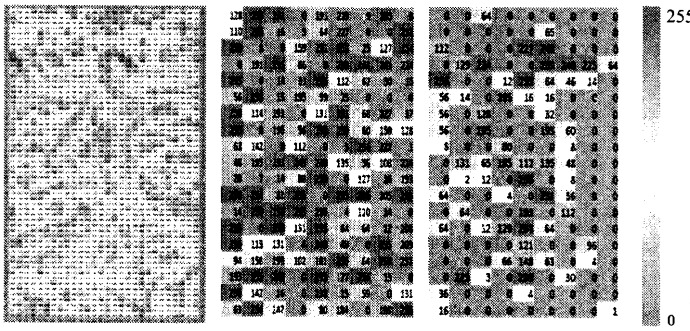
N: Total number of negative samples
Var_i: Gradient variance of image i defined by

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

W, H: Width and height of the gradient image, respectively.
$\overline{Gr_i}$: Average gradient of image i defined by

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

This choice may seem trivial or even crude. However, since our approach is solely based on pedestrian shape extraction, the goal behind it is to actually set a value of 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.



(a) (b) (c)

Fig.3. (a) Original gradient image. (b) LBP image with τ =0, (c) LBP image with τ =Var.

As shown in **Fig. 3**, the use of the classical threshold (τ=0) does not provide a very accurate LBP image. By accurate we mean that the uniformly cluttered background (grass/trees) still produces noisy patterns and interferes with the desired shape information we are seeking to extract. On the other hand, using τ=Var provides a more accurate shape description.

V.The classifier

A non-linear support vector machine is trained and used for the classification (Libsvm[36]). We use a radial basis function kernel (RBF). It is unarguably clear that, throughout the literature [36], it has been shown that non-linear SVMs outperform linear ones, even though the main disadvantage of non-linear SVMs is the training time. 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 support vector machines are basically a constrained optimization problem that consists of minimizing the following quantity:

$$\min_{w,b,\xi}\left(\frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i\right)$$

Subject to:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$$

w : *Normal Vec. to the separating hyperplane*
C : *Penalty term for erronous Classification*
ξ : *Distance to the hyperplane of misclassified points*

Here the kernel $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is equal to:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

The RBF kernel $K(x_i, x_j)$ maps each vector of the training data to a point in a higher dimensional space. It uses nonlinear separators but, within the kernel space, it constructs a linear equation. The support vector machine attempts to separate the points x_i's into subsets with homogeneous target values y_i's.

VI. Training and results with discussion

As specified in the introduction, the approach presented in this paper is tested on the INRIA pedestrian dataset [11], which offers a good benchmark dataset with high variations of poses and backgrounds.

The original INRIA dataset consists of a 1,208 pedestrian images (with their mirrored reflections) training set and a test set of 288 images with 589 human samples (also with their reflections) and 453 human free images. For the training data, the original image size is 96x160 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 30x60 pixels.

Since the INRIA dataset is an unbalanced set, evaluating just the accuracy does not give a good insight about the performance. This is why, in this paper, for each given configuration, we choose the SVM parameters (C, γ) to maximize the area under the curve (AUC) of receiver operating characteristic (ROC). Moreover the parameters that maximize the AUC with 5-fold cross validation were chosen each time to prevent over-fitting.

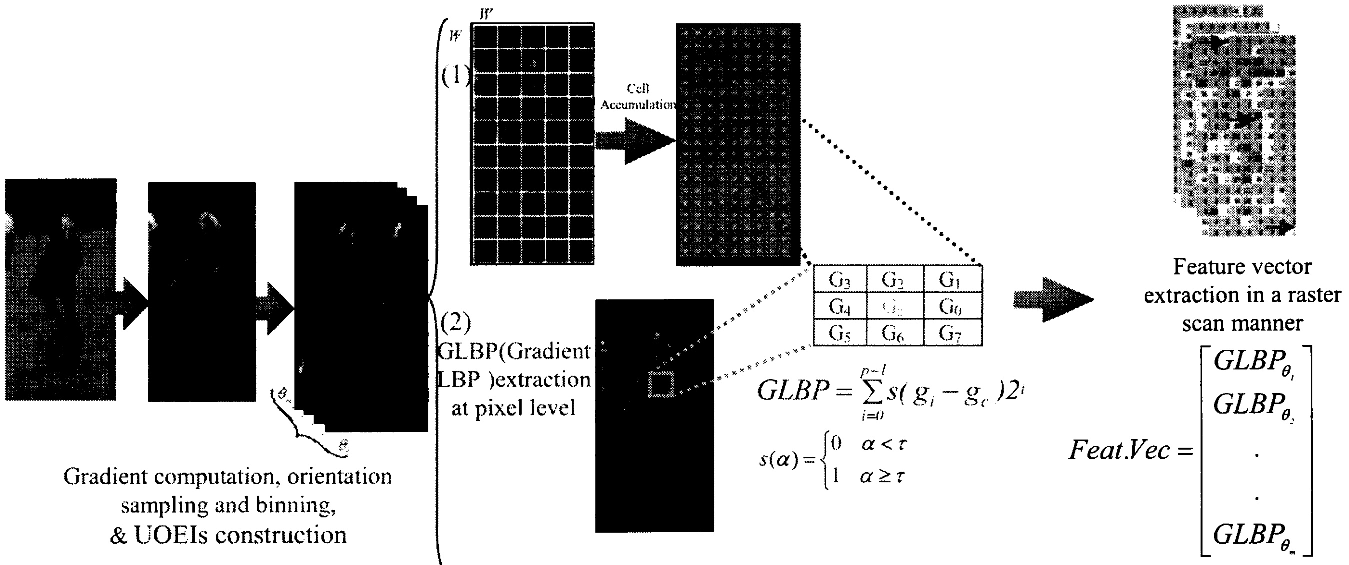

W
(1)
Cell Accumulation
G3 G2 G1
G4 G0
G5 G6 G7
(2)
GLBP(Gradient LBP)extraction at pixel level
$GLBP=\sum_{i=0}^{p-1}s(g_i-g_c)2^i$
$s(\alpha)=\begin{cases}0 & \alpha<\tau\\ 1 & \alpha\geq\tau\end{cases}$
Feature vector extraction in a raster scan manner
$Feat.Vec=\begin{bmatrix}GLBP_{\theta_1}\\ GLBP_{\theta_2}\\ \cdot\\ \cdot\\ GLBP_{\theta_m}\end{bmatrix}$
θ
Gradient computation, orientation sampling and binning, & UOEIs construction

Fig .4. Overview of the approach: (1) reproducing the method introduced in [5] by 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

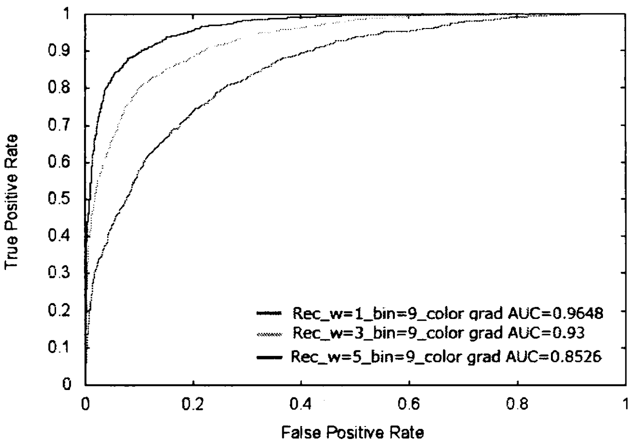
In [5], the patterns of oriented edge magnitude applied to face recognition show that the best performance is achieved with the number of bins $m=3$, which is understandable since, in their application, the main characteristic of a feature vector is discriminability. On the other hand, in human detection,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

All the results that are presented in the following employ unsigned gradient orientation with $m=9$.

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

1- Window size w= {3, 5, 7, 9}:

[REC= Rectangular LBP Neighborhood
w= Accumulation window size: w=1 means no accumulation
Bin= bin number m
Gray/Color_grad: We tested the effect of using color gradient as compared to gray scale gradient.



From the graph above, we see that the larger the accumulation window is, the less is the system 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 mid-low resolution images.

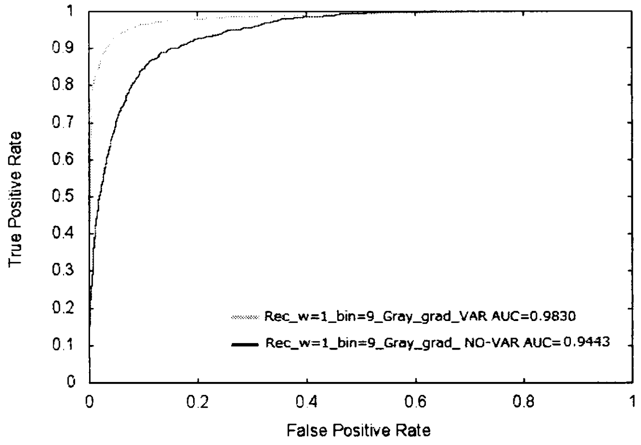
This provides us with the main advantage that the LBP operator can be applied directly and the accumulation step is skipped as well as the integral image computation step. This is already a gain in the computation time.

2- Classical threshold(τ=0) vs. new threshold (τ=Var)

This choice represents the main originality that this study brings to the field. The previously depicted results were for a classic threshold τ =0(*NO-Var*) [25].

We can see clearly on the following graphic 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. This result confirms clearly the conjecture we stated earlier.

[*Var= taking the threshold τ as equal to the average of the variance of negative samples*
NO-VAR= using the classical LBP threshold τ =0]



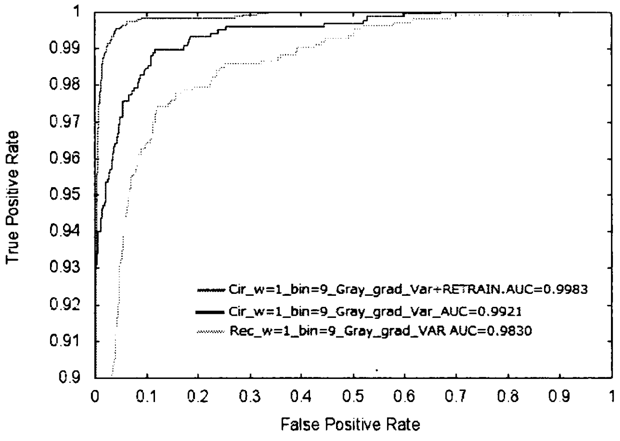
To improve efficiency, and getting the best performance possible of this approach, we also tried the circular LBP block geometry, making use of the bilinear interpolation techniques when the desired pixel value does not lie on the center of a pixel. We noticed a slight improvement of the performance.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

Also, to minimize the false positives, we used what is called retraining, employed in the original HOG paper [1] . This technique consists of training a model with the available negative/positive samples, and then we took 10,000 random 30x60 pixel patches from the negative images, we use them as test data, and retrain the original model by adding the false positives that have been generated. By adding what is called the "hard examples" to the training model, the recall improved by 6%.

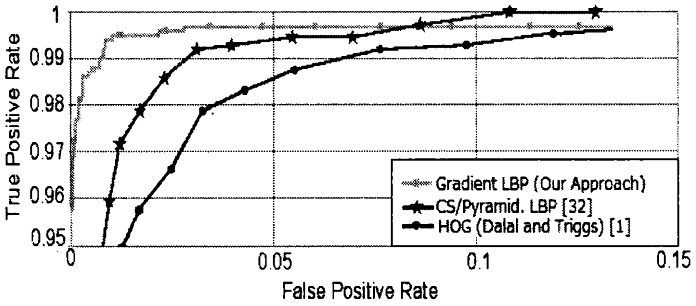
Here below, we depict both the effect of applying the circular neighborhoods and the retraining.

[*Cir= Circular neighborhood using bilinear interpolation Retrain= adding the false positives to the negative set and retraining the model*]

For the highest curve, with a circular neighborhood, and with orientations sampled to 9 directions, using gray scale gradient, combined with the threshold τ $=Var$ and with retraining, we obtained an accuracy of 98.44% at 10^{-2} False positives per window(FPPW). On a Core 2 Duo 3.00 GHz processor with 2Gb of memory, the time it took to extract the feature vector was 239 micro seconds.



In the next graph, 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 outperforms a more recent pedestrian detection framework tested on the INRIA pedestrian dataset, which is the center symmetrical and pyramidal LBP.

Putting aside the classification-wise comparison, the strongest point of this approach is the simplicity.

VII. Conclusions and future work

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

Also a GPU implementation is planned aside with an occlusion handling scheme.

Acknowledgments

This study was supported by JSPS KAKENHI under Grant-in-Aid for scientific Research (22510177)

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR 2005, volume 1, pages 886–893, 2005.

[2] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615–1630, 2005.

[3] M. Enzweiler and M. Gavrila. Monocular pedestrian detection: Survey and Experiments, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.31 n.12, p.2179-2195, December 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] Ngoc-Son Vu, Alice Caplier. Face recognition with patterns of oriented edge magnitudes. In ECCV 2010. pp. 313-326

[6] P. Doll'ar, B. Babenko, S. Belongie, P. Perona, and Z. Tu. Multiple component learning for object detection. In ECCV, 2008.

[7]Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: ICCV. Volume 1. (2005) 786–791 Vol. 1.

[8] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 4, pp. 349-361, Apr.2001.

[9] A. Ess, B. Leibe, K. Schindler, and L. van Gool. A mobile vision system for robust multi-person tracking. In CVPR, 2008.

[10]Zhang, B., Shan, S.S., Chen, X., Gao: Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Transactions On Image Processing 16 (2007) 57–68

[11] N. Dalal, "Finding People in Images and Videos," PhD thesis, Institut Nat'l Polytechnique de Grenoble, 2006.

[12]Crow, F. (1984). Summed-area tables for texture mapping. SIGGRAPH, 84, 207–212.

[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'l J. Computer Vision, vol. 38, pp. 15-33, 2000.

[15] E. Seemann, M. Fritz, and B. Schiele. Towards robust pedestrian detection in crowded image sequences. In CVPR, 2007.

[16] D. Tran and D. Forsyth. Configuration estimates improve pedestrian finding. In NIPS, volume 20, 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 Computer Vision and Pattern Recognition, 2007.

[19] A. Shashua, Y. Gdalyahu, and G. Hayun. Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In Intelligent Vehicles Symposium, 2004.

[20] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," Int'l J. Computer Vision, vol. 63, no. 2, pp. 153-161, 2005.

[21] E. Seemann, M. Fritz, and B. Schiele, "Towards robust pedestrian detection in crowded image sequences," IEEE Computer Vision and Pattern Recognition, 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 Transactions on Pattern Analysis and machine intelligence, 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 Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1475-1490, Nov. 2004.

[27] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In Proc. ECCV'04, pp.469–481, 2005.

[28] XY Wang, X. Han and SC Yan. An HOG-LBP human detector with partial occlusion handling. In ICCV, 2009.

[29]T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1996) pp. 51-59.

[30] D.G. Lowe, "Distinctive image features from scale invariant keypoints" Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[32] 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), pages 281–292, 2011.

[33] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou. Discriminative local binary patterns for human detection in personal album. In *CVPR*, 2008

[34] Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: Real-Life Images Workshop at ECCV. (2008)

[35] http://www.cse.oulu.fi/MVG/LBP_Bibliography .

[36] Chang, Chih-chung, and Chih-jen Lin. "LIBSVM: a library for support vector machines." *Science* 2.3 (2001) : 1-39.