NON-REDUNDANT GRADIENT SEMANTIC LOCAL BINARY PATTERNS FOR PEDESTRIAN DETECTION

Jiu Xu, Ning Jiang and Satoshi Goto

Graduate School of Information Production, and System LSI, Waseda University, Japan
xujiu@ruri.waseda.jp; jiangning@ruri.waseda.jp; goto@waseda.jp

ABSTRACT

In this paper, a feature named Non-Redundant Gradient Semantic Local Binary Patterns (NRGSLBP) is proposed for pedestrian detection as a modified version of conventional Semantic Local Binary Patterns (SLBP). Calculations of this feature are carried out for both intensity and gradient magnitude image so that texture and gradient information are combined. Moreover, non-redundant patterns are adopted on SLBP for the first time, allowing better discrimination. Compared with SLBP, no additional cost of the feature dimensions NRGSLBP is necessary and the calculation complexity is considerably smaller than that of other features. Experimental results on several datasets show that the detection rate of our proposed feature outperforms those of other features such as Histogram of Orientated Gradient (HOG), Histogram of Templates (HOT), Bidirectional Local Template Patterns (BLTP), Gradient Local Binary Patterns (GLBP), SLBP and Covariance matrix (COV).

Index Terms**—**Pedestrian detection, feature extraction, non-redundant gradient semantic local binary patterns

1. INTRODUCTION

Human detection is becoming a focusing research area in computer vision. It can be an important preliminary step for content analysis topics, such as behavior recognition, human tracking and etc.. Pedestrian detection, in particular, is widely applied in intelligent vehicle and visual surveillance systems. The primary aim of pedestrian detection is to locate the positions of human forms by using bounding boxes to distinguish them from the background in captured images.

Presently, two representative approaches can be concluded for human detection. One is called part based methods [1] and another is sub-window based methods [2-7].

In the sub-windowing method, an invariably sized rectangular sliding window is used to scan an input image from the top left to the bottom right in different scale spaces densely. Then, for each sliding window, features are extracted and input to a classifier that is trained offline in advance through labeled training samples. The classifier returns a value to determine whether a given window contains a human or not by comparing the feature vectors inside this sliding window with the data from the classifier.

In part based human detection methods, human body is divided into several parts and detectors for each part are trained separately. The training and detection parts for each detector are similar to the process of the sub-window based method. After the results of each part is achieved, combination and merging method are applied to generate the global result.

In this paper, a novel feature called Non-Redundant Gradient Semantic Local Binary Patterns is defined to achieve high detection rate for pedestrian detection. The proposed feature is a combination and modification of Semantic Local Binary Patterns (SLBP) [6] and Non-Redundant Local Binary Patterns (NRLBP) [8]. It should be mentioned that our proposal also belongs to this sub-window based method.

The rest of this paper is organized as follows. Section 2 explains related work. In Section 3, proposed NRGSLBP feature is described in detail. Section 4 presents the experimental results, and finally, conclusion is summarized in Section 5.

2. RELATED WORK

Feature extraction is an important topic in pedestrian detection fields. Over the past decade, many promising methods have appeared and these features can be mainly divided into two groups: gradient-based features and texture-based feature.

The Histogram of Orientated Gradient (HOG) [2] and the Covariance matrix (COV) [4] are two representative gradient-based approaches. In particularly HOG, proposed by Dalal and Triggs, is a benchmarking algorithm for pedestrian detection. It can achieve excellent performance in describing the human shape through a gradient direction histogram of small pieces using gradient information.

Some texture-based features, such as LBP-based method [6,9], also contribute great results to the pedestrian detection. The most attractive advantages of LBP are its invariance to

Thanks to the members of Research and Development Center, TOSHIBA Corporation for helpful discussions during internship.

monotonic intensity changes, low computational complexity and convenient multi-scale extension. SLBP and Fourier-LBP (FLBP) [6] have been proposed as effective applications of conventional LBP to the problem of human detection.

Recently, a number of features have involved a combination or modification of multiple complementary features on the basis of the above mentioned texture-based and gradient-based features. For example, NRLBP was used in [8] by providing a more compact description of human appearance. It reflected the relative contrast between the background and foreground and achieved better performance. Tang proposed a Histogram of Template (HOT) in [7] by combining the gradient and textural features, that used eight templates and four formulas to extract the local template-based binary features. Bidirectional Local Template Patterns (BLTP) [10] for human detection were proposed based on HOT and Center-symmetric Local Binary Patterns (CSLBP) [11]. The BLTP calculated several formulas for two directions of each template, which could be more discriminative than unidirectional one. In [12], Gradient Local Binary Patterns (GLBP) was proposed by adding the gradient magnitude value into corresponding 2-D table similar to SLBP.

Unlike these methods, we apply the non-redundant patterns on conventional SLBP together with feature extraction on both intensity and gradient magnitude map so that the texture and gradient information are combined without additional length of feature vectors. The proposed method and experimental results on benchmark data are described in the following sections.

3. PROPOSED FEATURE

3.1. Overview of Semantic Local Binary Patterns

The original $LBP_{8,1}$ requires 256-bin histogram for each 3 × 3 blocks, although longer feature vector is more sparse and discriminative, it is significantly costly in terms of memory storage and classification complexity. Moreover, the decimal codes of the original LBP have a large distance when the patterns are rotated by a certain degree. Fig. 1 is an example of this. Fig. 1(a) shows a binary code of 11110000, whereas Fig. 1(b) shows a binary code of 01111000. These two structures are similar for describing humans and should be loaded into spatially close histogram bins. However, the decimal codes of the two patterns have a significant difference: the former is 240 and the latter is 120.

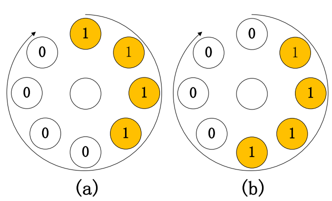

Fig. 1. Example of the limitation of LBP in describing similar structures for pedestrian detection

SLBP is proposed in [6] as a modified version of conventional LBP to solve the aforementioned problem. The calculation of this feature requires two steps. First, this feature calculates two parameters: angle and length from each LBP code. In the next stage, the corresponding histogram bins are voted according to the angle and length to generate a 2-D histogram.

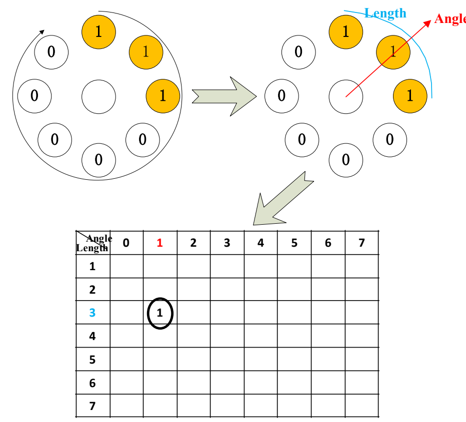

Fig. 2. Feature extraction process of SLBP for each pixel

Fig. 2 shows the principle process. The authors use the number of “1” pixels as the length and the position of the middle pixel in these continuous “1” pixels as the angle from the binary code of uniform $LBP_{8,1}$ to extract a 2-dimensional histogram. Then, all the pixels vote same weight “1” to their bins of in this 2-D histogram. For each block, the authors transform the bin values of 2-D histogram to a 1-D vector for Support Vector Machine (SVM) training.

Although the SLBP can arrange similar feature into connected histogram bins and address the problems of noise and illumination variations, limitations still remain. The first limitation is that SLBP is sensitive to the relative changes between the background and the foreground, although it is extremely efficient in orientation classification. For example, in Fig. 3, the SLBP codes for the enlarged 3 × 3 regions of curves A and B are completely different: curve A has a length of three and an angle of five, whereas curve B has a length of five and an angle of two. However, both curves actually represent the same structure in the human detection aspect which means the SLBP has redundant histogram bins in its 56-bin length histogram.

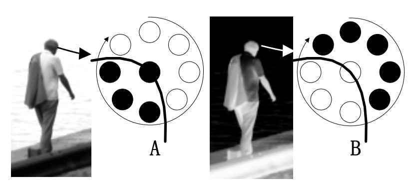

Fig. 3. Limitations of SLBP in human detection

Another limitation of SLBP is the loss of gradient information. In the HOG feature, the gradient values of pixels are voted to the histogram as the weights, but in SLBP a fixed weight are voted only on the intensity map only. Pixels with large magnitude values are always crucial for classification. Thus, SLBP cannot obtain enough information from these important pixels with larger magnitude values.

3.2. Non-Redundant Gradient Semantic Local Binary Patterns

To mitigate the limitations of conventional SLBP, we propose the NRGSLBP feature. The feature extraction of this NRGSLBP contains four steps.

In step 1, the input sub-window is divided into several blocks. For each pixel inside one block, the length and angle of its $LBP_{8,1}$ code are calculated as the conventional SLBP. Fig. 4 shows the look-up table for the coding: the numbers inside the table are the decimal values of $LBP_{8,1}$ code. According to the $LBP_{8,1}$ code of each pixel, the corresponding angle and length can be accessed.

Angle / Length	0	1	2	3	4	5	6	7
1	128	64	32	16	8	4	2	1
2	129	192	96	48	24	12	6	3
3	193	224	112	56	28	14	7	131
4	195	225	240	120	60	30	15	135
5	227	241	248	124	62	31	143	199
6	231	243	249	252	126	63	159	207
7	247	251	253	254	127	191	223	239

Fig. 4. Look-up table for NRGSLBP step 1

In step 2, non-redundant patterns are applied to each pixel. For each pixel, if the decimal value of its $LBP_{8,1}$ code is greater than 127, the original $LBP_{8,1}$ is replaced by its complementary value. As is shown in Fig. 5, the values in red in Fig. 5(a) are replaced by the values in blue in Fig. 5(b). The sum of red and blue value in corresponding position is 2^8-1. For example, when the angle of a pixel's $LBP_{8,1}$ is two and its length is four, the decimal value of this code is 240 which is greater than 127. Thus, the decimal value is replaced by 15 and the sum of 240 and 15 is 2^8-1. Therefore, all decimal values that match, as shown in Fig. 5(b), where the values from the four columns on the left match the values from the four columns on the right, are voted to the same histogram bin. Note that the total bin length is reduced from 56 to 28. By utilizing this framework, curve A and cure B in Fig. 3 are treated as the same structure which makes the feature more robust and discriminative.

In step 3, to compensate the loss of the gradient information, a first order derivative operator is used on the intensity image to obtain the gradient magnitude map. Then, non-redundant patterns are adopted on this map and the calculation is the same as in step 2.

In the last step, for each block in the image, a 28-bin length histogram on intensity map and a 28-bin length histogram on gradient magnitude map, both of which are generated from step 2 and step 3, respectively, are combined together as the proposed 56-bin length NRGSLBP. Finally, 56-bin length histograms of each block inside the image are summarized as the final feature set.

Angle / Length	0	1	2	3	4	5	6	7
1	128	64	32	16	8	4	2	1
2	129	192	96	48	24	12	6	3
3	193	224	112	56	28	14	7	131
4	195	225	240	120	60	30	15	135
5	227	241	248	124	62	31	143	199
6	231	243	249	252	126	63	159	207
7	247	251	253	254	127	191	223	239

(a)

Angle / Length	0	1	2	3	4	5	6	7
1	127	64	32	16	8	4	2	1
2	126	63	96	48	24	12	6	3
3	62	31	112	56	28	14	7	124
4	60	30	15	120	60	30	15	120
5	28	14	7	124	62	31	112	56
6	24	12	6	3	126	63	96	48
7	8	4	2	1	127	64	32	16

(b)

Same bin

Same bin

Fig. 5. Non-redundant patterns are applied for NRGSLBP step 2

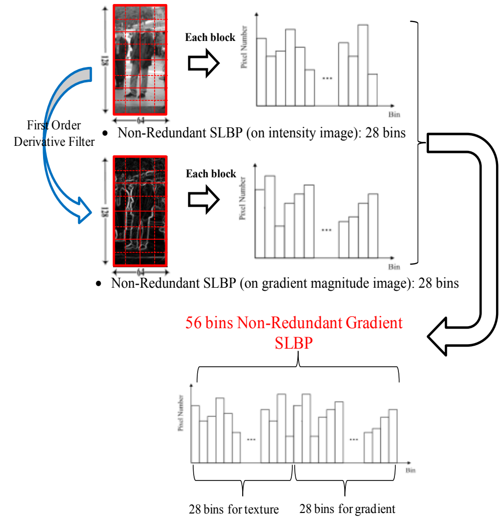

Fig. 6. Non-redundant patterns are calculated on both intensity and gradient magnitude maps

Fig. 6 illustrates the main feature extraction process of our proposal. By associating texture with gradient information in feature extraction, the proposed algorithm calculates more data on the human body for each training sample than other features like HOG and COV. Compared with SLBP, no additional cost of the feature dimensions NRGSLBP are necessary and the calculation complexity of this LBP-based approach is considerably low. Furthermore, by combining SLBP with NRLBP, the proposed NRGSLBP inherits the advantages of each and improves upon both. NRGSLBP is not only robust in the relative contrast between the background and foreground, but it also arrange similar pixel into connected histogram bins. It should be noticed that, to the best of our knowledge, this is the first time that non-redundant patterns are applied on SLBP. All these factors make it more discriminative than its predecessor.

4. EXPERIMENTAL RESULTS

4.1. Training method

The choice of classifiers is important, and it significantly affects the detection result. In order to get a fair comparison with other methods, in our experiments, SVM [13] is selected to train the detector and LibSVM [14] are used as tools. We allow all training parameters retain their default settings. A linear kernel function was used because its training and classification speeds are higher than those of nonlinear kernels. The decision rule is given by the following formula:

$$f(x) = \sum_{i=1}^{N_s} \beta_i K(x_i, x) + b \quad (1)$$

$$\text{Linear kernel: } K(x, y) = x' \times y \quad (2)$$

where each x_i is supported by the support vector, β_i is its weight, N_s is the number of support vectors, and $K(x, y)$ is the kernel function.

4.2. Evaluation method

We plot the Detection Error Trade-off (DET) curves for evaluation which is a representative method used in other studies [2]. A DET curve reveals how detection rate (miss rate) changes with the rate of the false positive per window (FPPW) curve. The terms are defined as follows:

$$\text{Miss rate} = 1\text{- recall} = \frac{\text{Number of false detection in positive samples}}{\text{Total number of positive samples}} \quad (3)$$

$$\text{FPPW} = \frac{\text{Number of false detection in negative samples wimdows}}{\text{Total number of negative windows}} \quad (4)$$

4.3. Benchmark datasets

We evaluate our feature performance on two datasets. The first evaluation is on an INRIA [15] datasets, which is widely used for human detection, it helped us to make fair comparisons with other methods. The dataset contains 1774 human annotations and 1671 human free images. Specifically, 1208 human annotations and 1218 human-free images are used for training the detectors, and the remaining images are used for testing.

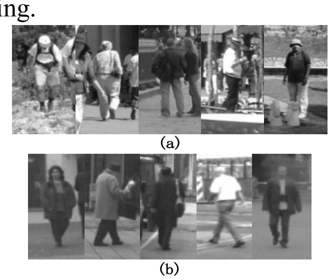

Fig. 7. Positive training samples (a) INRIA dataset (b) Daimler dataset

The second dataset called Daimler Mono Pedestrian Detection Benchmark Dataset [16], is used for evaluation. This dataset is made from industrial cameras. It is different from the previous mentioned INRIA dataset which is captured by consumer cameras. For the training part, this dataset contains 15660 positive patches and 6742 negative natural images captured from vehicle views. 21792 sequential images with pedestrian annotations are utilized for testing. Some samples in INRIA and Daimler dataset are shown in Fig.7.

4.4. Performance comparisons

We compared our proposed NRGSLBP feature with the HOG [2], COV [4], SLBP [6], HOT [7], BLTP [10], and GLBP [12]. Both the feature extraction framework and training dataset were the same; thus, the comparison is unbiased. As with the HOG, we selected negative samples in the resampling stage and we applied a normalization method. In this experiment, we chose the NRGSLBP feature with a block size of 16 and stride of 8, so the length of our feature for a 64 × 128 image was 5880. The result of HOG, HOT, SLBP, BLTP and GLBP were also implemented by using a block size of 16 and stride of 8 with linear kernel function while COV used variable block sizes.

Fig. 8 shows the overall performance comparison on INRIA dataset, the last curve is our proposal. Here "Lin" denotes a linear kernel function; thus, the SVM training methods are the same. The results shown in Fig. 8 prove that our proposed method achieves the best performance in comparison with other related method: our method could achieve a 90.1% detection rate at 10^{-5} FPPW and a 95.5% detection rate at 10^{-4} FPPW.

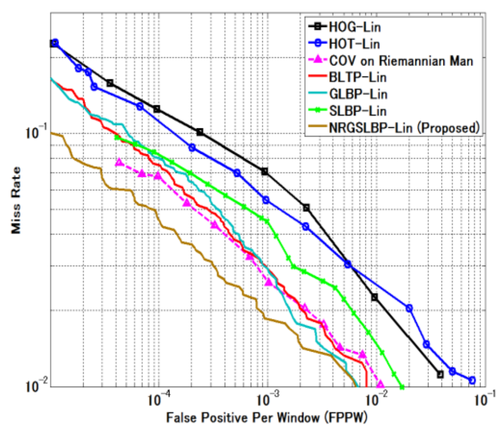

Fig. 8. Performance comparisons on INRIA dataset

We also evaluated our feature using the Daimler dataset and compared it with the other related features. All the detectors were trained through linear SVM for unbiased comparisons. Fig. 9 shows that the performance of NRGSLBP is the best among all methods. The detection rates shown in Fig. 8 and Fig. 9 prove that our proposed feature achieved better results than the other referred features on multiple datasets. Furthermore, the results indicate that the proposed feature is suitable for images composed from both industrial and consumer cameras.

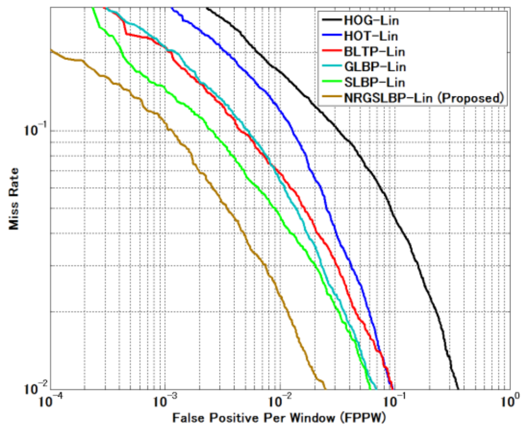

Fig. 9. Performance comparisons on Daimler dataset

In order to emphasize the contribution of our proposal, detection rates of each stage in section 3.2 are given in Fig. 10, the results were evaluated by INRIA dataset, still trained by linear SVM. By applying the Non-Redundant framework as mentioned in step 2, the performance of the 28-bin Non-Redundant SLBP is better than conventional 56-bin SLBP from 10^{-3} FPPW to 10^{-2} FPPW. So the overall performance is improved, but not by a wide margin. However, the feature length is reduced by half and short length represents smaller workload and memory cost. Moreover, after the combination of intensity and gradient information performed in step 3 and 4, the result can be further improved significantly.

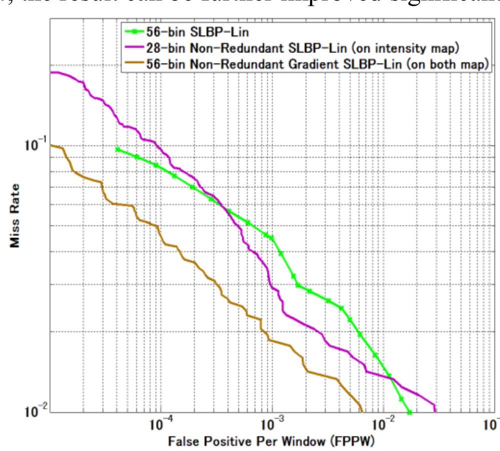

Fig. 10. Performance comparisons of each stage on INRIA dataset

5. CONCLUSION

In this paper, a robust human detection method is given. A novel powerful local feature is defined as NRGSLBP as a modified version of conventional SLBP. In this feature, non-redundant patterns are adopted on SLBP initially, allowing better discrimination to the relative changes between the background and foreground. Moreover, feature extraction process are operated on both intensity and gradient magnitude maps in order to associate both texture and gradient information. Results on INRIA and Daimler datasets show that our NRGSLBP-based linear detector achieves the best detection rate compared with other linear detectors. Besides, the calculation complexity of proposed LBP based feature is considerably low and it is convenient for multi-scale extension. All these factors make our proposed NRGSLBP feature more promising than other referred features.

REFERENCES

[1] A. Hadid, M. Pietikäinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," Proc. of International Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 797–804, 2004.

[2] N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection", Proc. of International Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 886-893, 2005.

[3] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A Discriminatively Trained, Multi-Scale, Deformable Part Model", Proc. of International Conf. on Computer Vision and Pattern Recognition (CVPR), 2008.

[4] T. Oncel, and P. Fatih, "Pedestrian Detection via Classification on Riemannian Manifolds", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, No. 10, pp. 1713-1727, 2008.

[5] D. Lowe, "Distinctive image features from scale-invariant key points," International Journal of Computer Vision (IJCV), vol. 60, pp. 91-110, 2004.

[6] M. Yadong, and Y. Shuicheng, "Discriminative Local Binary Patterns for Human Detection in Personal Album", Proc. of International Conf. on Computer Vision and Pattern Recognition (CVPR), 2008.

[7] S. Tang, S. Goto, "Histogram of Template for Human Detection", Proc. of 35th International Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 2186-2189, 2010.

[8] D. T. Nguyen, Z. Zong, P. Ogunbona, and W. Li, "Object detection using non-redundant local binary patterns," Proc. of International Conference on Image Processing (ICIP), pp. 4609-4612, 2010.

[9] L. Nanni, and A. Lumini, "Ensemble of Multiple Pedestrian Representations", IEEE Trans. on Intelligent Transportation Systems, Vol. 9, No.2, 2008.

[10] J. Xu, N. Jiang, and S. Goto, "Pedestrian Detection Based on Bidirectional Local Template Patterns", Proc. of 20th European Signal Processing Conference (EUSIPCO), pp. 400-404, 2012.

[11] Y. Zheng, C. Shen, R.I. Hartley, and X. Huang, "Pyramid center-symmetric local binary/trinary patterns for effective pedestrian detection", Asian Conference on Computer Vision (ACCV), 2010.

[12] N. Jiang, J. Xu, W. Yu and S. Goto, "Gradient Local Binary Patterns for Human Detection", IEEE International Symposium on Circus and Systems (ISCAS), page 978-981, 2013.

[13] B. Scholkopf, and A. Smola, Learning with Kernels, Support Vector Machines, Regularization, Optimization and Beyond, MIT Press, 2002.

[14] LibSVM [Online], http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[15] INRIA Dataset [Online], http://lear.inrialpes.fr/data

[16] Daimler Mono Pedestrian Detection Benchmark Dataset [Online], http://www.gavrila.net/Datasets/datasets.html