Weighted HOG for Thermal Pedestrian Detection

[1]Yi-Hao Li, [1]Shih-Shinh Huang, [2]Ching-Hu Lu, and [3]Feng-Chia Chang

National Kaohsiung University of Science and Technology[1].
National Taiwan University of Science and Technology[2].
National Chung-Shan Institute Science and Technology[3]

***Abstract*-- Pedestrian detection is an important topic in many applications. For the purpose of applications used around the clock, the work for detecting pedestrian based on thermal sensors has attracted significant attentions. Based on the observation that the histograms of four cells in HOG descriptor over the pedestrian contour are significantly different, this study proposes a new feature called weighted HOG (WHOG) that weights the traditional HOG by the difference pattern of the histograms of cells. In the experiment, four videos are used to validate the proposed WHOG and the results demonstrate that WHOG outperforms HOG in all cases.**

I. INTRODUCTION

Over the last decades, a significant amount of works on pedestrian detection using visible cameras has been done. However, the main limitation of pedestrian detection using visible cameras is its inapplicability in nighttime environment. Driven by the decreasing cost, far infrared (FIR) sensors relying purely on ambient heat patterns but not illumination source is naturally as the solution in detecting pedestrians at poor illumination conditions including nighttime, fog, heavy rains. Although various approaches have studied in recent years, developing a robust algorithm for pedestrian detection using a monocular FIR camera is still a challenging task due to the high variability in pedestrian appearance, the lack of texture information in thermal images, noisy background and occlusion effect.

To describe thermal pedestrian appearance, the direct way is with the use of representative templates. In [1][2], probabilistic templates capturing the variations in pedestrian shape are computed and further used for matching. The HOG (Histogram of Oriented Gradients) proposed by Dalal and Triggers [3] is an effective feature which has been widely used in both visible and thermal pedestrian detection. To improve the performance of the original HOG, several HOG variants have been proposed in the literature. Kim *et. al*. [4] proposes Histograms of Local Intensity Differences (HLID), which replaces the gradient magnitude and orientation of HOG feature by the maximal difference of eight neighboring points.

Although these features perform well in thermal pedestrian detection, they are all developed based on visible but not thermal images. Accordingly, the main objective of this study is to propose a new feature called weighted HOG which encodes the difference pattern into the HOG for thermal pedestrian detection.

This research is partially supported by the project grant MOST 106-2221-E-327 - 028

II. WEIGHTED HOG

By observing the HOG histogram of each cell as shown in Fig. 1, it is obviously that cell histograms inside the pedestrian (positive) interior and in the background (negative) scenes have similar distribution. The cell histograms of the blocks over the pedestrian contour are significantly different. According to this observation, the idea behind of WHOG is to weight the HOG by the difference pattern among the cells.

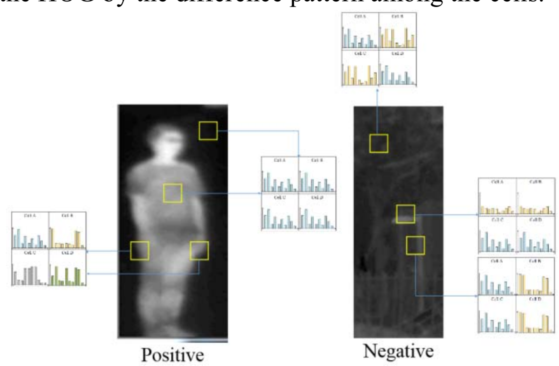


Fig. 1: An example of HOG histogram of several blocks in positive and negative samples.

Let B_i be a block and $C_i^{\{l,t\}}, C_i^{\{r,t\}}, C_i^{\{l,b\}}$, and $C_i^{\{r,t\}}$ are the HOG histograms of the four cells of B_i. Here (l, r, t, b) denote left, right, top, and bottom. We define the horizontal HOG difference $d_h(B_i)$ and vertical one $d_v(B_i)$ of cells in B_i as:

$$d_h(B_i) = \sum_{j=1}^{9} \left| C_i^{l,t}(j) - C_i^{r,t}(j) \right| + \sum_{j=1}^{9} \left| C_i^{l,b}(j) - C_i^{r,b}(j) \right|$$

$$d_v(B_i) = \sum_{j=1}^{9} \left| C_i^{l,t}(j) - C_i^{l,b}(j) \right| + \sum_{j=1}^{9} \left| C_i^{r,t}(j) - C_i^{r,b}(j) \right| \quad (1)$$

where $C_i^{\{l,t\}}(j)$ is the jth bin of the histogram of left-top cell. As a consequence, the HOG difference in the block B_i is defined as:

$$\mathrm{d}(B_i) = d_h(B_i) + d_v(B_i) \quad (2)$$

Thus, the HOG difference is used to weight the HOG so that to improve its discriminability. The proposed WHOG feature is thus defined as (3)

$$\mathrm{WHOG}(B_i) = \mathrm{w}(B_i) \times \mathrm{HOG}(B_i) \quad (3)$$

where $w(B_i)$ is the weight from the HOG difference and its definition is expressed as the following equation:

$$\mathrm{w}(B_i) = \frac{d(B_i)}{\sum_{j=1}^{9}\left(c_i^{l,t}(j) + c_i^{l,b}(j) + c_i^{r,t}(j) + c_i^{r,b}(j)\right)} \quad (4)$$

The denominator term of $w(B_i)$ is used for normalization. $\text{HOG}(B_i)$ is the HOG feature descriptor of the block B_i

III. EXPERIMENT

In this section, the scene for video collection to validate the performance of the proposed method is in NKFUST campus. The infrared thermal camera used in our experiments is a FLIR Tau 2 which is mounted on a tripod. The frames of the thermal videos are 14-bit gray-level images with resolution 640x512. The dataset for training is from 11 persons, 3 scenes, 2 time periods (daytime and nighttime), 2 clothing styles, 2 poses, 4 distances (30m, 40m, 50m, 60m), and 8 view angles. This results in a dataset consisting of 11x3x2x2x2x4x8=8448 pedestrian images. Some examples of two persons in 8 view angles are shown in Fig. 2. The negative samples are cropped from the images without pedestrians and number of negatives used in this study is also 8448.

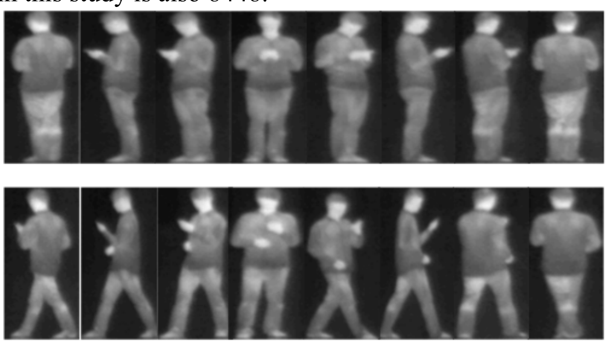
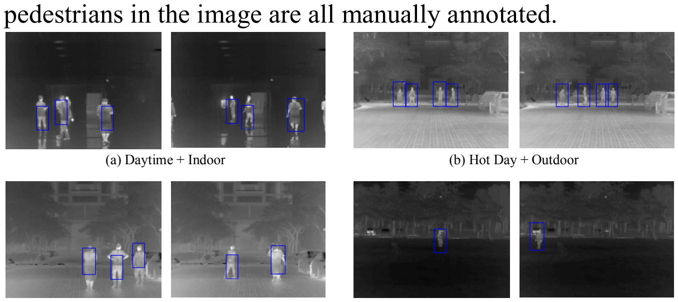


Fig. 2: Examples of positive samples

To validate the learned pedestrian classifier using WHOG, we collect four videos in different scenarios including Daytime+Indoor, Hot Day + Outdoor, Nighttime+Outdoor, and Nighttime+Outdoor with complex background. The Hot Day + Outdoor scenario is to verify if the detector is effective in case of the background has similar temperature as human body; the Nighttime + Outdoor with complex background is for the case of the background with objects (i.e. tree) similar to pedestrians. Each video consists of 50 frames and the pedestrians in the image are all manually annotated.



(a) Daytime + Indoor (b) Hot Day + Outdoor

(c) Nighttime + Outdoor (d) Nighttime + Outdoor + Complex Background

Fig. 3. Detection results of four scenarios.

For validating the performance of the proposed WHOG, the curve of FPPI versus missing rate is used to evaluate the performance of the algorithm. The definition of matching is that the overlap of the detected window and the ground-truth one is over than 0.5. The overlap is expressed as follows:

$$a_0 = \frac{area(BB_{dt} \cap BB_{gt})}{area(BB_{dt} \cup BB_{gt})} > 0.5 \quad (5)$$

where BB_{dt} and BB_{gt} are detected window and the ground-truth one, respectively. Besides, we compare the proposed WHOG with the traditional HOG proposed in [3]. Fig 4. shows the resulting curves, respectively, for the four testing videos. Obviously, the proposed WHOG feature outperforms HOG in all cases, especially in Daytime+Indoor scenario. In Hot Day + Outdoor scenario, the SW-HOG is only slightly better than HOG. This is because the temperature of the ground plane is similar to human body so that the contour of pedestrian is not clear.

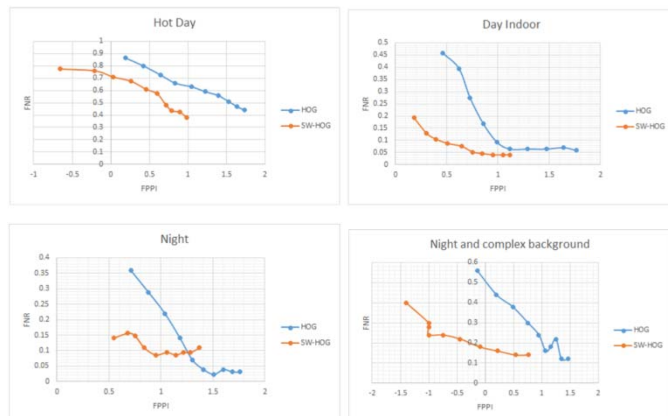


Fig. 4. FPPI (false positive per image) versus FNR (false negative rate) for four scenarios.

IV. CONCLUSION

Based on the analysis that the histograms of four cells in HOG descriptor over the pedestrian contour are significantly different, this study calculates the difference among the histograms of the cells in a block to weight the traditional HOG. The resulting feature is referred to the weighted HOG. The effectiveness of the proposed WHOG is validated by using four videos captured in NKFUST campus and with difference scenarios. According to the curves of FPPI versus missing rate, it demonstrates that WHOG outperforms the traditional HOG. In the near future, we should verify the weighted strategy to other well-known thermal features, such as HLID.

REFERENCE

[1] M. Bertozzi, A. Broggi, C. Hilario Gomez, R. I. Fedriga, G. Vezzoni, M. Del Rose, “Pedestrian Detection in Infrared Images based on the Use of Probabilistic Templates,” *IEEE Intl. Conf. on Intelligent Vehicles Symposium*, pp. 327—332, 2007.

[2] Harsh Nanda and Larry Davis, “Probabilistic Template Based Pedestrian Detection in Infrared Videos,” *IEEE Intl. Conf. on Intelligent Vehicle Symposium*, pp. 17—21, 2002.

[3] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 886–893, 2005.

[4] D. S. Kim, M. Kim, B. S. Kim and K. H. Lee, “Histograms of Local Intensity Differences for Pedestrian Detection in Far-Infrared Images,” *IEL Electronics Letters*, vol. 49, no. 4, pp. 258—260, 2013.