

A ROBUST IMAGE DESCRIPTOR FOR HUMAN DETECTION BASED ON HOG AND WEBER'S LAW

SHINFENG D. LIN*, YUAN-MING LIU AND YU-RUEI JHU

Department of Computer Science and Information Engineering
National Dong Hwa University

No. 1, Section 2, Dashiue Road, Hualien 97401, Taiwan

*Corresponding author: david@mail.ndhu.edu.tw

Received September 2012; revised January 2013

ABSTRACT. *Human detection is essential for many applications such as surveillance and smart car. However, detecting humans in images or videos is a challenging task because of the variable appearance and background clutter. These factors affect significantly human shape. In this article, a robust descriptor based on HoG and Weber's Law is proposed. Namely, the proposed descriptor is concatenated by U-HoG and histogram of Weber's constant. Weber's constant has advantages such as robust to noise and detecting edge well. Because there are a large number of weak edges in the cluttered background affecting the detection result, the proposed method uses Weber's constant to take off the weak edges. If a pixel is on the weak edge, the proposed method will ignore the pixel when computing the feature. Therefore, the proposed descriptor may inherit the advantages of Weber's Law. From the simulation results of human detection, it is noted that the key contribution of the proposed method is that it has better performance than other comparative methods and is more robust to Gaussian white noise than U-HoG.*

Keywords: Human detection, HoG, Weber's law, U-HoG, SVM, WLD

1. Introduction. Human detection is a very important task for a number of applications in computer vision, e.g., content-based image/video retrieval, smart vehicles, and surveillance systems. For example, human tracking and identification are highly dependent on reliable human detection. Some methods were proposed for detecting human face as Lu et al. [1] or human body as Zin et al. [2]. Although human detection has been well studied in vision, it still remains challenging in varying illumination conditions, including occlusion, background clutter, and noise influence. These factors affect significantly human shape and appearances. Therefore, how to develop a robust descriptor for human detection is essential. In recent years, people have focused on two research areas: looking for more discriminative descriptors or finding more powerful classifier.

Human detection using sliding window classifiers is presently a predominant method because of its good performance [3,4]. In the sliding window's approach, a rectangular sliding window densely scans from upper left to bottom right on each image in different scales. The image descriptors for human detection can roughly be classified into two categories depending on how the features are measured [5]. The first category operates on all images in all datasets. Principal component analysis is one of the well-known global descriptors. The disadvantage of global features is that the approach fails to extract significant features if there is a large variation in the object's appearance and illumination conditions. On the other hand, the second category operates on the subset regions of the images. The local descriptors are fewer sensitive to these problems, because the features are extracted from the subset regions of the images. Some instances of the famously used local descriptors are gradient orientation [4], wavelet coefficient [6], and

region covariance [7]. Mainly two types of local descriptor approaches are used, dense feature-based approaches and sparse feature-based approaches [8]. In the dense feature-based approaches, human shapes are modeled densely or globally over the shapes, e.g., an over-complete set for histograms of oriented gradients in [4], Haar wavelet features in [9], rectangular features in [10], locally deformable Markov models in [11], and covariance descriptors in [7]. Because of this characteristic, these approaches require a large number of training data.

Local descriptor approaches can further be divided into full body and body parts detection [5]. Several part detectors are trained separately for each body part in the part-based approach [12]. The individual results are combined with the second classifier to form the whole body detection. The part-based approach can deal with variation in human appearance due to body articulation. However, this approach increases the complexity of the human detection problem.

For the approach of densely feature-based full body detection, Dalal and Triggs [4] presented a human detection algorithm with excellent detection results. Their method uses a dense grid of Unsigned Histograms of Oriented Gradients (U-HoG) to represent a detection window. The Unsigned Histogram of Oriented Gradient descriptor provides excellent performance relative to other existing feature sets. However, U-HoG still has a problem that gradients of opposite directions in a cell are mapped to the same bin. In order to solve this problem, Satpathy et al. [13] proposed Ex-HoG which was concatenated by Diff-HoG and U-HoG. Also, another method based on HOG, Multislit-HOG, was proposed by Zin et al. [14]. The approaches of feature-based full body detection need to train a strong classifier. The classifiers in these approaches are much simpler than those of part body detection approaches, because the classifiers do not need to be trained separately.

The Support Vector Machine (SVM) is used to classify the windows into either human or non-human. The reasons why we select the SVM classifier [15] are as follows: 1) SVM is one of the famous and powerful classifier; 2) SVM classifier is easy to train and the global optimum is guaranteed. Thus, the problem caused by suboptimal training is avoided for fair comparison and 3) U-HoG and Ex-HoG used SVM. Therefore, we can demonstrate the performance of the proposed method by comparing with U-HoG and Ex-HoG when using the same classifier. In this article, we try to solve some problems in HoG and U-HoG for improving the performance of human detection. To achieve this goal, we propose a new descriptor based on Weber's Law and Unsigned Histogram of Oriented Gradient.

The rest of this article is organized as follows. We review background and related works in Section 2. The proposed human detection scheme is presented in Section 3. Section 4 demonstrates the experimental results. Finally, the conclusion is drawn in Section 5.

2. Background and Related Works. This section introduces the background of image descriptor in computer vision. Weber's Law Descriptor is briefly described in Section 2.1, the histogram of oriented gradients is introduced in Section 2.2, and the extended histogram of oriented gradients is addressed in Section 2.3.

2.1. Weber's Law descriptor. The Weber's Law descriptor was proposed by Chen et al. [16]. Weber's Law descriptor contains two components: differential excitation and orientation. Differential excitation is a function of the Weber fraction. Orientation is a gradient orientation of the current pixel. The flowchart of Weber's Law descriptor is illustrated in Figure 1.

2.2. Histogram of oriented gradients feature. The Histogram of Oriented Gradients was proposed by Dalal and Triggs [4]. It is another descriptor that uses gradient strength

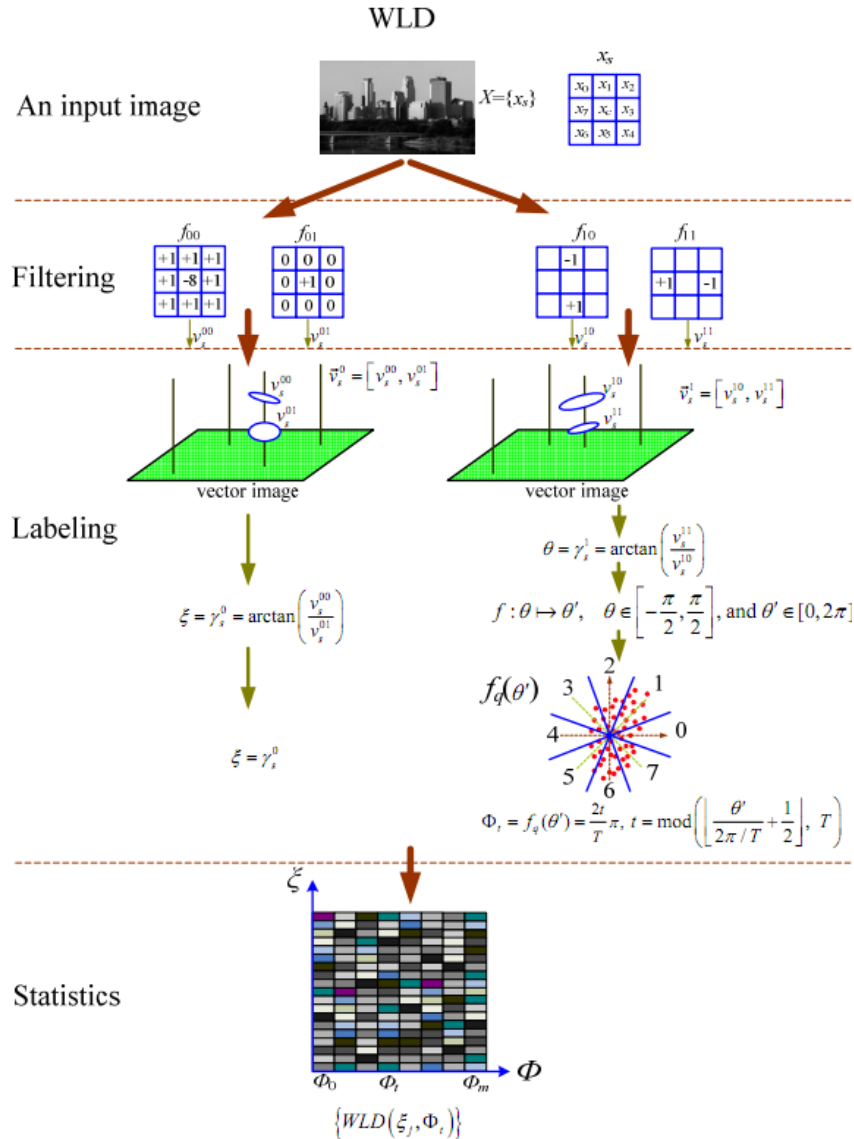


FIGURE 1. The flowchart of Weber's Law descriptor was proposed by [16]

and gradient orientation information for human detection. In this sub-section, HoG is presented, and the flowchart of HoG [17] is illustrated in Figure 2.

In the HoG approach, each sliding window is divided into non-overlapping cells. The descriptor blocks are densely computed and overlapped over uniformly sampled cells. Specifically, each block contains $\alpha \times \alpha$ cells, and each cell contains $\beta \times \beta$ pixels. The pixels in the cell are voted into γ bins histogram. Thus, a concatenated histogram of $\gamma \times \alpha^2$ is formed for each block. For the sliding window, the histograms from all the blocks are concatenated to form a feature vector.

2.3. Extended histogram of gradients feature. Satpathy et al. proposed the Ex-HoG feature in [13]. The Ex-HoG improves the performance of U-HoG for human detection because the Ex-HoG has success in avoiding the confusion with the U-HoG. Dalal and Triggs [4] proposed both Signed-HoG and Unsigned-HoG at the same time. Signed-HoG represents the situations whereby a bright object is against a dark background and a dark object is against a bright background in different feature vectors since it considers gradient directions from 0° to 360° . In the human detection problem, this makes the

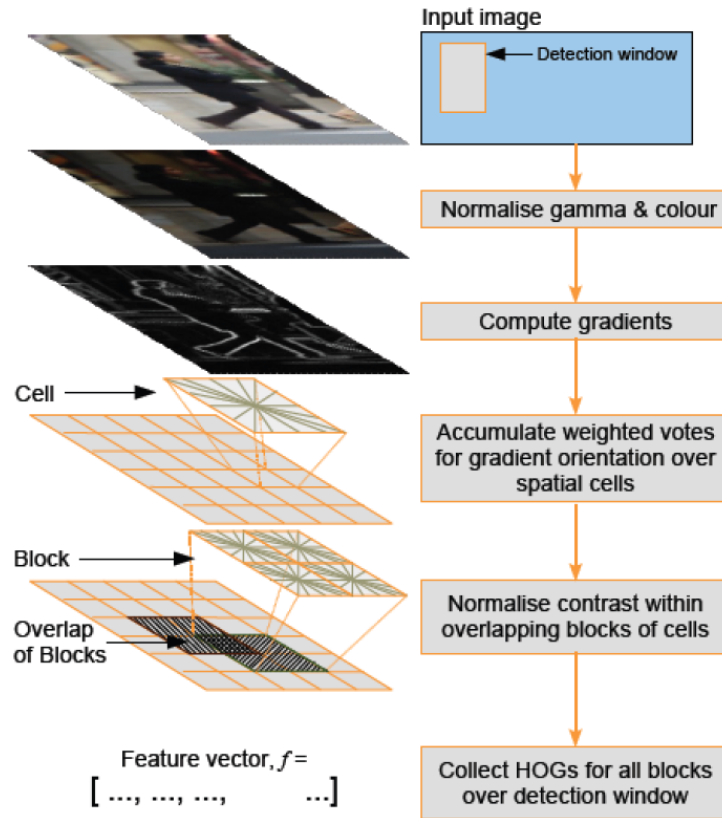


FIGURE 2. The flowchart of HOG algorithm was proposed by Jiang [17]

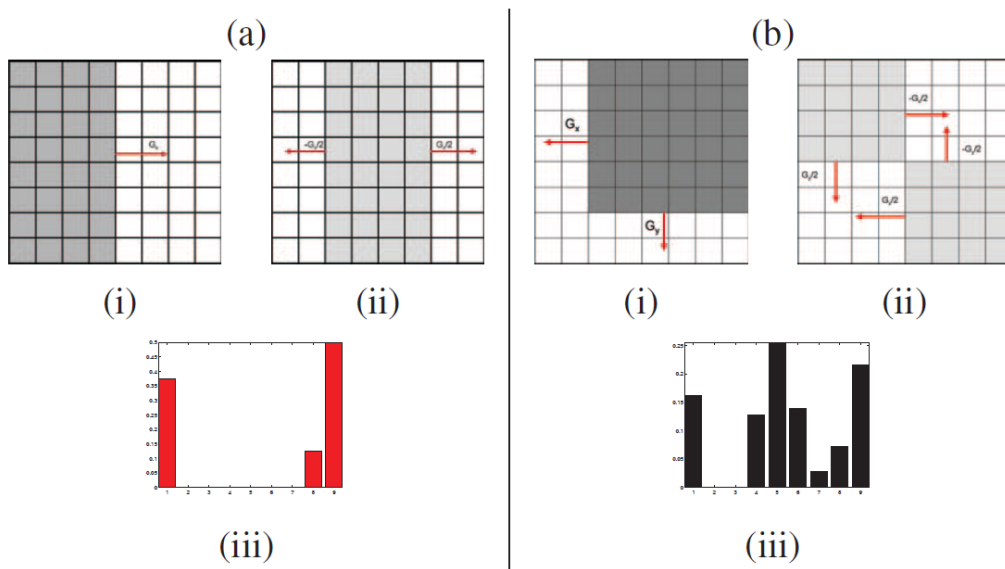


FIGURE 3. Some examples that different patterns are represented in similar feature vector when using U-HoG

intra-class variation of the human objects larger. To solve this problem, Dalal and Triggs proposed Unsigned Histogram of Gradients (U-HoG). The U-HoG only oriented the gradients of direction θ and $\theta + 180^\circ$ ($0^\circ < \theta < 180^\circ$) to be direction θ for alleviating the larger intra-class variation. However, this makes the fundamental problem with U-HoG that gradients of opposite directions in a cell are mapped to the same bin in a histogram [13]. Therefore, some patterns that are different from each other were unable to discern.

Hence, it is possible for two different patterns to have the similar feature vectors. Figure 3 shows an example of this problem. It confused the human detector on classifying human/non-human.

3. The Proposed Scheme. The general procedure of human detection can be roughly divided to two parts: feature extraction and feature classification. In this article, we focus on the step of feature extraction. The U-HoGs and the differential excitation in WLD are employed. Figure 4 illustrates the flowchart of the proposed scheme. The significance is to compete with Ex-HoG by combining U-HoG with Weber’s constant. Furthermore, the effect of weak edge is reduced by using Weber’s constant. In Section 3.1, we describe the motivation of the proposed scheme. From the figure, the red rectangle (Differential Excitation) is described in Section 3.2, and the green rectangle (Weak Edge Removal Algorithm and U-HoG) is described in Section 3.3. The procedure of feature collection is illustrated in Section 3.4.

3.1. Motivation of the proposed scheme. Satpathy et al. made a reference to U-HoG problem in the literature [13]. The problem was that gradients of opposite directions in a cell are mapped to the same bin in a histogram. This problem makes the different patterns be represented by the same feature vector. To solve this problem, Satpathy et al. proposed concatenating U-HoG and Diff-HoG to separate the different patterns represented when using U-HoG.

It is observed that if the different patterns are represented in similar feature vector when using U-HoG, the strength of edges are mutually different. For example, there are two different patterns A and B. If the number of pixels on each edge in the pattern A is twice as much as the number of pixels on each edge in the pattern B, and the strength of the edges in pattern A is half of the strength of the edges in pattern B. In addition, the directions of the half of edges in the patterns A and B are contrary, and the others are the same. Then, the patterns A and B will be represented in the similar feature vectors when using U-HoG. This problem is caused by U-HoG which uses the weighted magnitude for

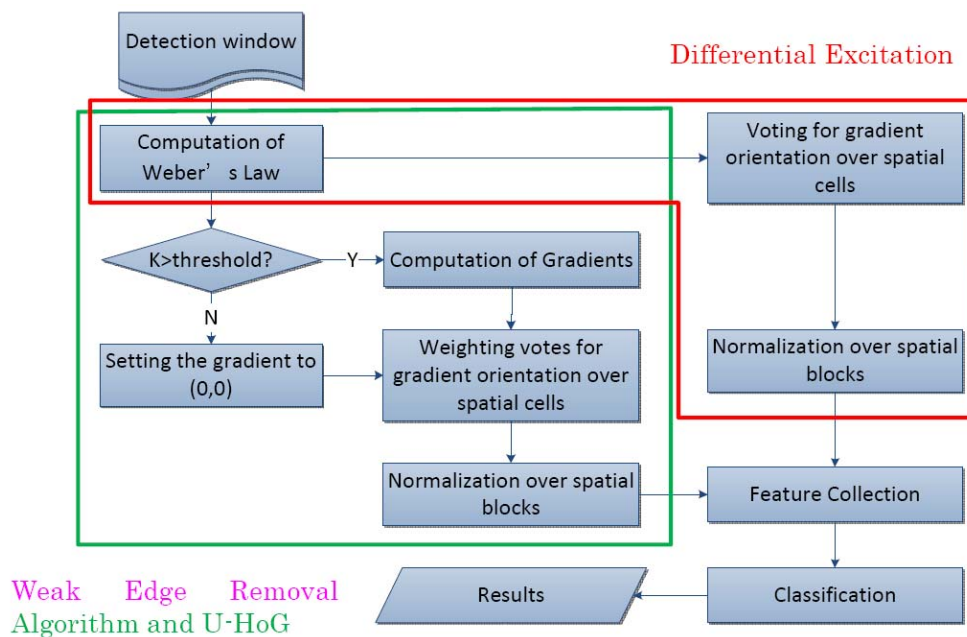


FIGURE 4. The flowchart of the proposed scheme

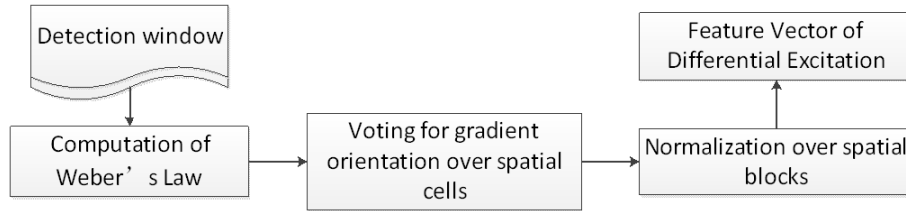


FIGURE 5. The flowchart of differential excitation

accumulating the histogram, and the contrary orientation of pixels would be voted to the same bin.

To further improve the performance of human detection and solve the problem when using U-HoG, we propose a new descriptor for human detection based on Weber's Law and U-HoG.

3.2. Differential excitation. To solve the problem of U-HoG, we apply the Weber's constant to assisting U-HoG. The Weber's constant can detect edges elegantly. Thus, the Weber's constant expresses the strength of edge well. The flowchart of differential excitation is illustrated in Figure 5.

In our approach, we do not use the method of accumulating the histogram in WLD. We refer HoG to accumulate the histogram. The size of the sliding window is 64×128 , each window is divided into non-overlapping 8×8 cells, and each block contains 2×2 cells. Each pixel in the cell calculates a vote into 9 bins according to the differential excitation of the pixel. Then, the histogram of blocks is concatenated by the histograms of cells in the block. So each histogram of block has 36 bins. And the proposed scheme uses the L1-norm, as (1), as a block normalization method. The inline-point similarity is computed as (5).

$$v' = \frac{v}{(\|v\|_1 + \varepsilon)} \quad (1)$$

3.3. Weak edge removal algorithm and U-HoG feature extraction. Human detection using sliding window classifiers is presently predominant method. In sliding window approaches, if the feature extraction method for human detection is based on gradients, it makes the accurate human detection difficult when the background is clutter. After observing a number of human images with clutter background, we perceive that clutter background usually has a large number of pixels on weak edge. These weak edges maybe affect the accuracy of human detection. Actually, Dalal and Triggs [4] tried to face this problem. They used the weighted magnitude for accumulating the histogram. This method let the importance of pixel on weak edge be lower. However, the U-HoG normalizes feature vector on block. If the edges in the block are all weak, the feature vector will be magnified. This makes the performance of human detection bad. In this article, we want to further reduce the effect of weak edges on detection rate. For this purpose, the proposed scheme employs Weber's constant in removing the weak edge from the image before extracting HoG feature. Because the detected edges using Weber's constant could elegantly match the subjective criterion [16], we believe that the Weber's constant can remove the weak edge well.

The flowchart of the weak edge removal algorithm and U-HoG is illustrated in Figure 6. We define a threshold to classify if the pixels are on edge. If a pixel is on edge, we compute the gradient of the pixel. On the other hand, if the pixel is not on edge, we ignore the pixel by setting the gradient of the pixel to $(0, 0)$. Thus, according to the algorithm of

U-HoG, the pixel will not be accumulated in the histogram. Then, we apply the typical U-HoG algorithm to computing the feature of sliding windows.

In our approach, we apply typical U-HoG to extracting feature. The setting details of the U-HoG are as follows [4]: We compute the gradients using the filters $[1 \ 0 \ -1]$ and $[1 \ 0 \ -1]^T$. The size of each sliding window is 64×128 , each window is divided into non-overlapping 8×8 cells, and each block contains 2×2 cells. Besides, the method of accumulating the histogram is also the same with typical U-HoG. The Gaussian weighted gradient magnitude of each pixel in the cell is voted into 9 bins according to the orientation of the pixel's gradients. Moreover, the orientation bins are evenly spaced over 0° - 180° (U-HoG). The votes are tri-linearly interpolated between neighboring bin centers in both orientation and position. Then, the proposed scheme uses L2-norm to normalize the vector on blocks.

3.4. Feature collection. After differential excitation and U-HoG feature extraction, we obtain two feature vectors from each block in the sliding window. As described in

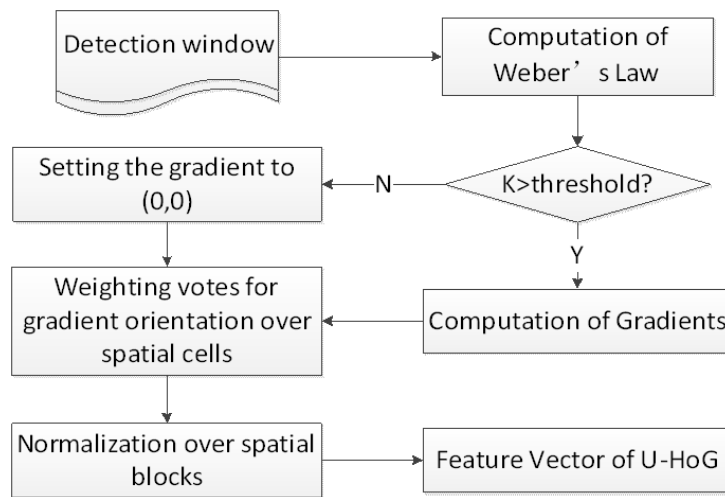


FIGURE 6. The flowchart of the weak edge removal algorithm and U-HoG

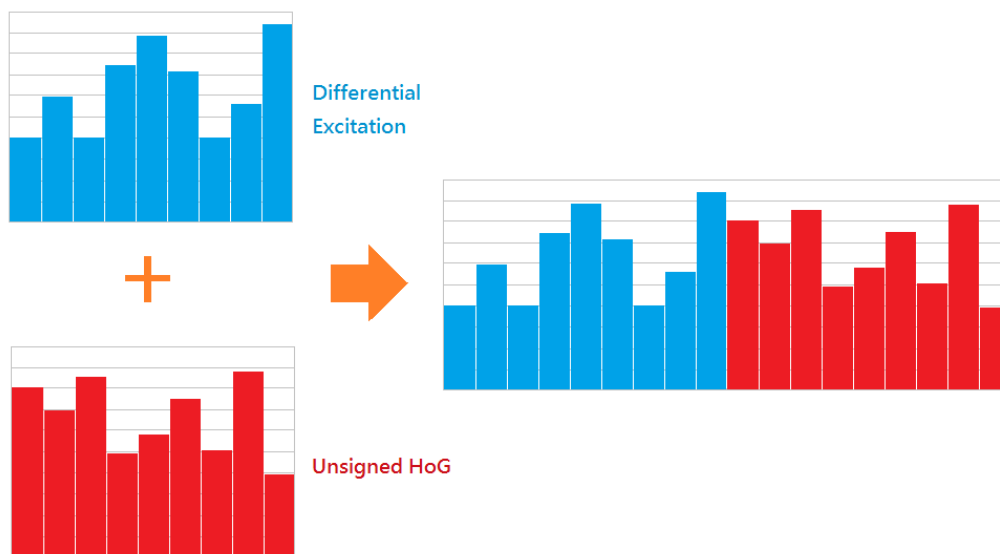


FIGURE 7. The proposed descriptor is concatenated by U-HoG and differential excitation in a cell

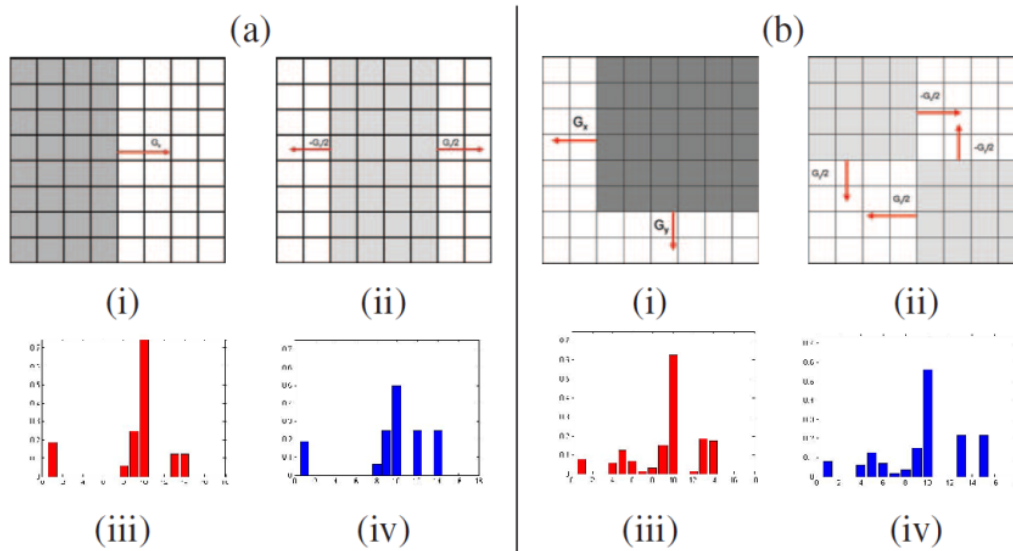


FIGURE 8. The proposed representations of patterns in Figure 3

Sections 3.1 and 3.2, each sliding window is divided into 105 overlapping blocks, and each block contains four cells. In addition, we obtain a 9 bins histogram from each cell. So, both U-HoG and differential excitation extract a 36D (9×4) feature vector from each block. Then, the proposed descriptor is concatenated by U-HoG and differential excitation (Figure 7). The final feature vector of each sliding window is concatenated by blocks in the window. Finally, each window is treated as a 7560D ($105 \times 36 \times 2$) feature vector. Our representations of patterns in Figure 3 are shown in Figure 8. From this figure, it is noted that the proposed method can also solve the problem of U-HoG.

4. Experimental Results. We examine the proposed descriptor with a famous database and the simulation results are shown as follows. Section 4.1 introduces the INRIA person database. In Section 4.2, we compare the results of the proposed method with existing methods. Experiments with adding Gaussian white noise are designed in Section 4.3.

4.1. INRIA person database. Dalal and Triggs [4] showed that U-HoG outperformed all other feature extraction methods before it. Recently, Satpathy et al. [13] showed that the performance of Ex-HoG was better than U-HoG. Hence, we test the performance of the proposed feature against U-HoG, HoG, and Ex-HoG using the INRIA Person Database [4]. The INRIA person database contains both training and testing sets. The training sets contain positive and negative samples of size 64×128 pixels and negative images. In addition, the positive training set contains 2416 human samples (1218 human samples and their mirrored samples) of fixed size, and the negative training set consists of 1218 non-human images. On the other hand, the testing sets also contain both positive and negative sets. The positive testing set consists of 1126 positive samples, and the negative testing set contains 453 non-human images. Some training and testing human samples from INRIA person database [4] are shown in Table 1.

4.2. Comparison with other methods. We use the same training and testing procedure as the methods in Dalal and Triggs [4] and Satpathy et al. [13]. When training the classifier, 10 negative samples are taken randomly from each negative training image. Therefore, there are 12180 negative training samples in the training set for the first round training. After first round training, the 1218 images are scanned densely to find false positive samples. These false positive samples are named ‘hard samples’. Then, retraining

TABLE 1. Some training and testing human samples from INRIA person database [4]



is performed using the augmented set (initial 12180+hard samples) to produce the final classifier.

Figure 9 shows the miss rate (1-recall) against the false positives per window (FPPW) for the compared methods. The lower curve means that the performance is better. In the human detection problem, the researchers often use miss rate at FPPW 10^{-4} , the fourth dotted vertical line in Figure 9, as a reference point for detection results. From Figure 9, the performance of U-HoG is better than HoG, because the U-HoG reduces the intra-class variation of human when using HoG. However, U-HoG maps the opposite directions into the same histogram bin. It makes the different patterns be described in similar feature vectors. Ex-HoG was proposed to solve this problem. So, the performance of Ex-HoG is better than U-HoG. However, our approach is proposed to compete with Ex-HoG. In our approach, the problem of U-HoG is solved by using Weber's constant. The effect of weak edge is reduced by using Weber's constant as well. The advantage of the proposed

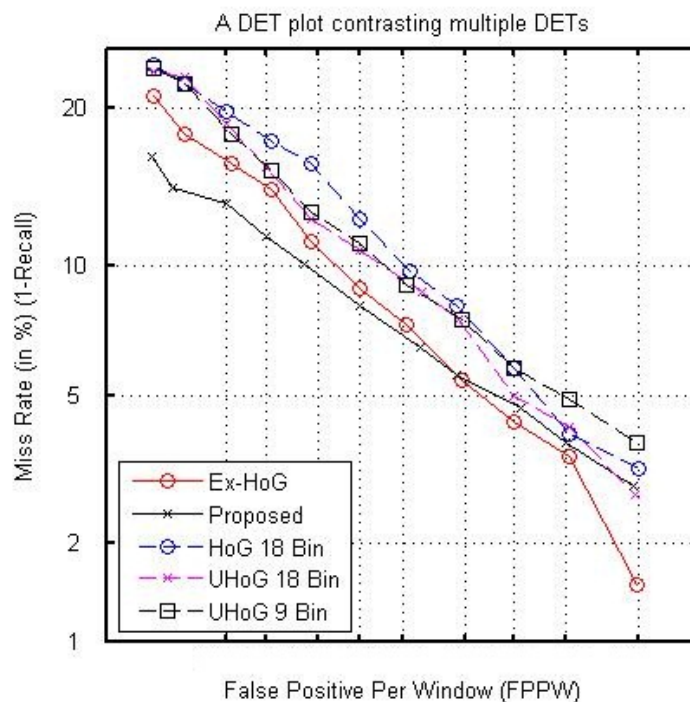


FIGURE 9. Performance comparison among the proposed method, HoG, and U-HoG in INRIA person database



FIGURE 10. Some detection results of the INRIA person database

method is illustrated in Figure 9. In Figure 9, the performance of the proposed method is better than Ex-HoG at $FPPW 10^{-4}$. Specifically, at $FPPW 10^{-4}$, the proposed method has a miss rate of 8.18% compared with the miss rate of 8.97% for Ex-HoG, 10.92% for U-HoG (9 Bins), and 11.9% for HoG.

In the proposed method, we use a rectangular sliding window to densely scan from upper left to bottom right on each image at different scales in the image pyramid. Scale ratio between two consecutive levels in the pyramid is 1.2 or 1.05. Figure 10 shows some detection results of the INRIA person database. Figure 11 and Figure 12 demonstrate some comparisons between the proposed method and U-HoG at $FPPW 10^{-4}$ using the same scale ratio. In the figures, the green rectangle is produced by U-HoG, and the red rectangle is produced by the proposed method. In Figure 12, while U-HoG misses one of the targets, this result demonstrates the effectiveness of the proposed method.



FIGURE 11. Comparison between the proposed method and U-HoG at FPPW 10^{-4}

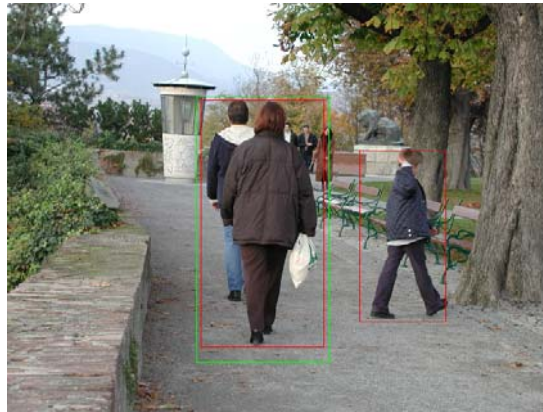


FIGURE 12. Comparison between the proposed method and U-HoG at FPPW 10^{-4}

4.3. Comparison with HoG after adding Gaussian white noise. In the literature [16], the differential excitation (Weber's constant) has advantages like robustness to noise and detecting edges. Specifically, a Weber's constant reduces the effect of noise, as it is similar to smoothing the image. Because the differential excitation (Weber's constant) is computed by a sum of its neighbor differences to the current pixel, differential excitation reduces the effect of noise pixels. Moreover, the sum of its neighbor differences is further divided by the intensity of the current pixel, which also decreases the effect of noise in an image.

In this sub-section, it is demonstrated that the Weber's constant could assist the U-HoG in resisting the effect of the noise. First, we detect human in the images in INRIA person database with added white Gaussian noise. The function `imnoise()` in Matlab is applied to adding Gaussian white noise. Table 2 shows some examples of the Gaussian white noise effect. In Table 2, the strength of noise is correlating with standard deviation (σ) in the function. To detect human in these images, we use the model which we have trained. And we also try to find out the detection rate at FPPW 10^{-4} . Then, we compare the

TABLE 2. Some examples of the Gaussian white noise effect

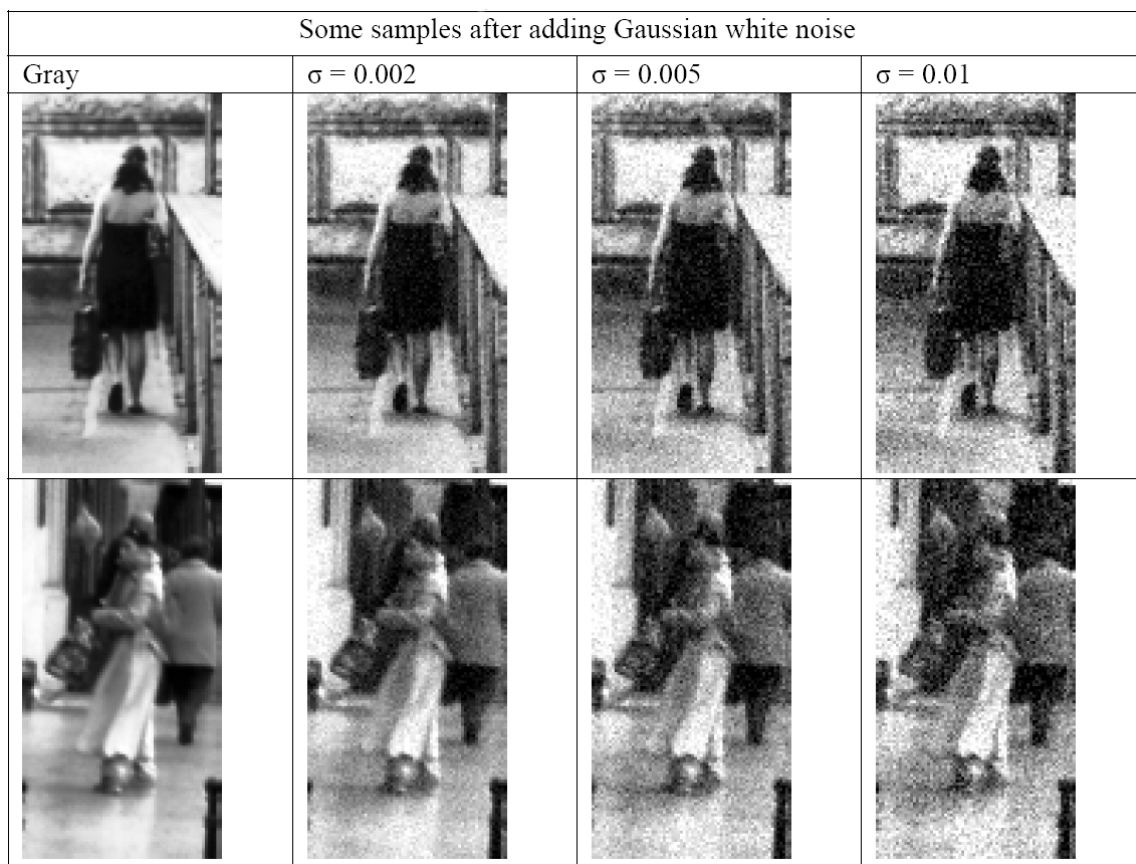






TABLE 3. A positive sample correctly classified using the proposed method but misclassified by U-HoG after adding Gaussian white noise

				
	$\sigma = 0$	$\sigma = 0.001$	$\sigma = 0.005$	$\sigma = 0.01$
HoG	√	√		
Proposed method	√	√	√	√

performance of the proposed descriptor with U-HoG. Figure 13 shows the performance at FPPW 10^{-4} . In this figure, the x axis is standard deviation (σ), and the y axis is the detection rate. The higher the curve, the better is the performance. We can see that the proposed descriptor is more stable than U-HoG in Figure 13. Therefore, there are some samples correctly classified using the proposed method but misclassified by U-HoG

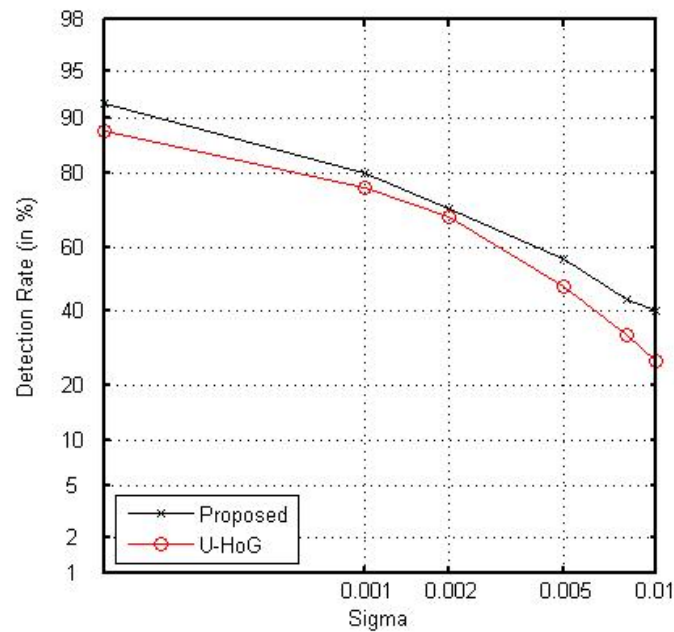


FIGURE 13. Comparison of the proposed descriptor against U-HoG at FPPW 10^{-4}

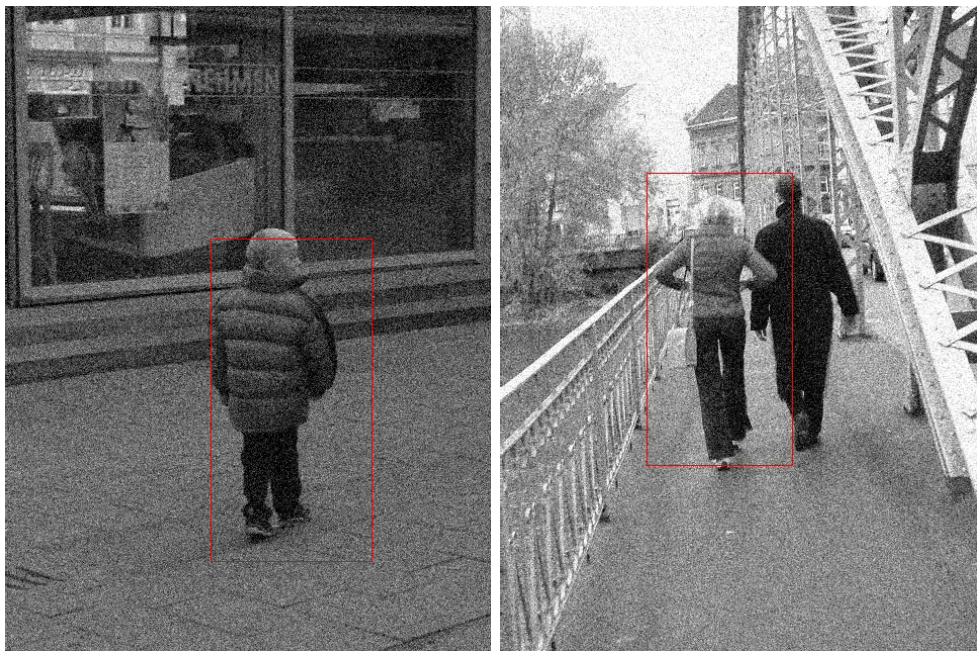


FIGURE 14. Comparison between the proposed method and U-HoG after adding Gaussian noise ($\sigma = 0.01$)

after adding Gaussian white noise. An example is shown in Table 3. This represents the significant contribution of the proposed method. Figure 14 presents some comparisons between the proposed method and U-HoG after adding Gaussian white noise ($\sigma = 0.01$).

The practical applications of the proposed method are addressed in this paragraph. Due to monitor screen usually has a lot of noise, we then try to detect human in the images which are captured from monitor screen. To do this, we download some images from Internet, and these images are captured from monitor screen. In Figure 15, Figure



FIGURE 15. Comparison between the proposed method and U-HoG

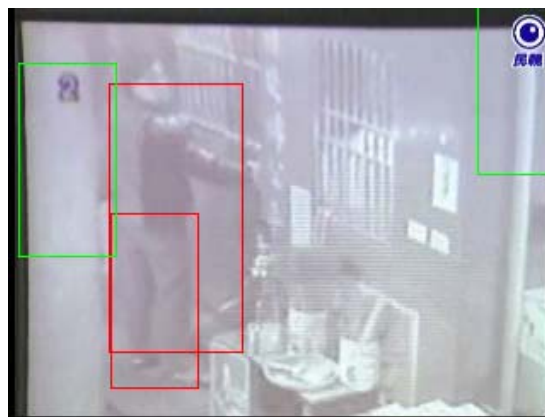


FIGURE 16. Comparison between the proposed method and U-HoG



FIGURE 17. Comparison between the proposed method and U-HoG

16 and Figure 17, it is observed that the images are blurred. However, the proposed method still can find the human correctly.

In these results, we can see that the proposed method is more robust than U-HoG, although U-HoG is not too weak to resist the Gaussian white noise. We believe this is because U-HoG uses gradient and the histogram computation. On the other hand, the results show that Weber's constant can assist U-HoG in resisting the Gaussian white noise.

5. **Conclusion.** A robust descriptor for human detection based on HoG and Weber's Law has been proposed. First, Weber's constant is extracted before U-HoG's extraction. Then, it adopts a threshold to check the pixels whether the pixels are on strong edge or not. Second, histogram of Weber's constant is accumulated, similar to U-HoG's accumulating the gradient. Third, the proposed feature is concatenated by U-HoG and differential excitation. Finally, a linear SVM is applied to classifying the features.

From the experimental results, the proposed descriptor has a miss rate of 8.18% compared with the miss rates of 10.92% for U-HoG and 8.97% for Ex-HoG at FPPW 10^{-4} . And the experimental results show that the proposed descriptor is more robust than U-HoG when adding white Gaussian noise.

REFERENCES

- [1] Z.-M. Lu, X.-N. Xu and J.-S. Pan, Face detection based on vector quantization in color images, *International Journal of Innovative Computing, Information and Control*, vol.2, no.3, pp.667-672, 2006.
- [2] T. T. Zin, H. Takahashi and H. Hama, Robust person detection in far infrared images – Methods based on multi-slits and GC movement patterns, *International Journal of Innovative Computing, Information and Control*, vol.5, no.3, pp.751-761, 2009.
- [3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, *The PASCAL Visual Object Classes Challenge*, 2008.
- [4] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of Conf. CVPR*, vol.1, pp.886-893, 2005.
- [5] S. Paisitkriangkrai, C. Shen and J. Zhang, Fast pedestrian detection using a cascade of boosted covariance features, *IEEE Trans. on PAMI*, vol.18, no.8, pp.1140-1151, 2008.
- [6] P. Viola and M. J. Jones, Robust real-time face detection, *Int. Journal of Computer Vision*, vol.57, pp.137-154, 2004.
- [7] O. Tuzel, F. Porikli and P. Meer, Human detection via classification on Riemannian manifolds, *Proc. of Conf. CVPR*, Minneapolis, MN, USA, pp.1-8, 2007.
- [8] Z. Lin and L. S. Davis, Shape-based human detection and segmentation via hierarchical part-template matching, *IEEE Trans. on PAMI*, vol.32, no.4, pp.604-618, 2010.
- [9] C. Papageorgiou, T. Evgeniou and T. Poggio, A trainable pedestrian detection system, *Proc. of Symp. Intelligent Vehicles*, pp.241-246, 1998.
- [10] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. of Conf. CVPR*, vol.1, pp.511-518, 2001.
- [11] Y. Wu, T. Yu and G. Hua, A statistical field model for pedestrian detection, *Proc. of Conf. CVPR*, vol.1, pp.1023-1030, 2005.
- [12] A. Mohan, C. Papageorgiou and T. Poggio, Example-based object detection in images by components, *IEEE Trans. on PAMI*, vol.23, no.4, pp.349-361, 2001.
- [13] A. Satpathy, X. Jiang and H. L. Eng, Extended histogram of gradients feature for human detection, *Int. Conf. on Image Processing*, Hong Kong, pp.3473-3476, 2010.
- [14] T. T. Zin, P. Tin and H. Hama, Pedestrian detection based on hybrid features using near infrared images, *International Journal of Innovative Computing, Information and Control*, vol.7, no.8, pp.5015-5025, 2011.
- [15] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. on Intelligent Systems and Technology*, vol.2, no.3, 2011.
- [16] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen and W. Gao, WLD: A robust local image descriptor, *IEEE Trans. on PAMI*, vol.32, no.9, pp.1705-1720, 2010.
- [17] W. Jiang, Human feature extraction in VS image using HOG algorithm, *Research Project*, University of Science & Technology of China, 2007.