

# Gabor-enhanced histogram of oriented gradients for human presence detection applied in aerial monitoring



Anton Louise De Ocampo <sup>a,1,\*</sup>, Argel Bandala <sup>a,2</sup>, Elmer Dadios <sup>b,3</sup>

<sup>a</sup> De La Salle University – ECE Department, 2401 Taft Ave, Malate, Manila 1004, Philippines

<sup>b</sup> De La Salle University – MEM Department, 2401 Taft Ave, Malate, Manila 1004, Philippines

<sup>1</sup> [anton\\_louise\\_deocampo@dlsu.edu.ph](mailto:anton_louise_deocampo@dlsu.edu.ph); <sup>2</sup> [argel.bandala@dlsu.edu.ph](mailto:argel.bandala@dlsu.edu.ph); <sup>3</sup> [elmer.dadios@dlsu.edu.ph](mailto:elmer.dadios@dlsu.edu.ph)

\* corresponding author

## ARTICLE INFO

### Article history

Received June 22, 2020

Revised July 8, 2020

Accepted July 9, 2020

Available online November 12, 2020

### Keywords

Human Detection

Gabor Filters

Local Image Patterns

Aerial Monitoring

Surveillance

## ABSTRACT

In UAV-based human detection, the extraction and selection of the feature vector are one of the critical tasks to ensure the optimal performance of the detection system. Although UAV cameras capture high-resolution images, human figures' relative size renders persons at very low resolution and contrast. Feature descriptors that can adequately discriminate between local symmetrical patterns in a low-contrast image may improve a human figures' detection in vegetative environments. Such a descriptor is proposed and presented in this paper. Initially, the acquired images are fed to a digital processor in a ground station where the human detection algorithm is performed. Part of the human detection algorithm is the GeHOG feature extraction, where a bank of Gabor filters is used to generate textured images from the original. The local energy for each cell of the Gabor images is calculated to identify the dominant orientations. The bins of conventional HOG are enhanced based on the dominant orientation index and the accumulated local energy in Gabor images. To measure the performance of the proposed features, Gabor-enhanced HOG (GeHOG) and other two recent improvements to HOG, Histogram of Edge Oriented Gradients (HEOG) and Improved HOG (ImHOG), are used for human detection on INRIA dataset and a custom dataset of farmers working in fields captured via unmanned aerial vehicle. The proposed feature descriptor significantly improved human detection and performed better than recent improvements in conventional HOG. Using GeHOG improved the precision of human detection to 98.23% in the INRIA dataset. The proposed feature can significantly improve human detection applied in surveillance systems, especially in vegetative environments.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## 1. Introduction

Human detection has been used in many applications such as crime mitigation [1], surveillance [2] [3], crowd-estimate or people counting [4], abnormal event recognition [5], person reidentification [6], gender classification [7], elderly fall detection [8] [9], illegal trespassing detection [10], and rescue operations [11]. In applications of human detection where the subject video or images are captured using unmanned aerial vehicles (UAV), the difficulty of automatic human detection is amplified by the variations in pose, the angle, and the distance by which a camera is capturing an image or video sequence mounted in UAV. Furthermore, the area which includes any human figure is relatively small compared to the extent of the image rendering the Region-of-Interest (ROI) at low resolution.

Vision-based human detection systems are generally equipped with visible light, FIR, NIR, IR cameras, or a combination of any of these. Each system differs in the features used to characterize human

figures. Some human detection algorithms are based on the subject's silhouette or edges [12], matching templates [13], sector rings [14], thermal images [15], visible human attributes [16], spatiotemporal features [17][18], covariance descriptor [6], and feature maps derived from convolutional neural networks [2][19]. Histogram of Oriented Gradients (HOG) is used for pedestrian detection [20], which performed relatively well for upright human figures. Since then, HOG has been considered by most researchers due to its robustness against varying illumination. Improvements in HOG are often made by supplementing it with another feature descriptor. In Hurney *et al.* [21], HOG is used in conjunction with Local Binary Patterns LBP to detect human figures during the night.

Similarly, LBP and HOG features are extracted from images containing human figures, then subjecting them to Singular-Value Decomposition and Principal Component Analysis PCA, to select the most significant feature segments from LBP and HOG [22]. Human detection in uncontrolled environments can also be done by using a textured image instead of the original. This is accomplished by preprocessing an image with Gabor filtering then extracting HOG features afterward [23]. Similarly, HOG features and Haar wavelets are combined to form a cascade head-shoulder detector that works well for overhead and tilted captures from surveillance systems [24]. Another hybrid use of HOG with LSS has been proposed to improve the feature descriptors' description capability for human detection [25].

Although there are various descriptors proposed, automatic human detection is still a challenging task. Holistic human detection methods perform well in varying illumination, but the performance degrades under occlusions or pose variations [26]. As mentioned above, among the other feature descriptors, HOG has been considered by most researchers as standard for several years. However, conventional HOG features cannot detect human figures in the low-contrast region since gradient computation is less effective or when local patterns occurred symmetrically. Most recent improvements on HOG, which are proposed to address these issues, include Histogram of Edge Oriented Gradients (HEOG) [26] and Improved HOG [27]. Other literature claims that combined features are superior to the use of any individual descriptor alone [28]; however, such a system becomes exclusive to a specific dataset [29]. It is in this context that this paper is proposed.

The main contribution of this paper is a proposal for a new feature computation for human detection. This paper proposes and evaluates the Gabor-enhanced HOG feature descriptor to address the issue of discrimination of local patterns at low-contrast and with symmetrical occurrence. Finally, the paper constructs a dataset intended to detect humans in the vegetative background at very low resolution.

## 2. Method

In human detection, two major steps are performed: first, extraction of the Region-of-Interest (ROI), and second, validation of the human target in the proposed ROI. The proposed method falls under the second step in which the feature extraction of the candidate is generated. As previously mentioned, HOG has been a persisted standard in holistic human detection. Hence, before detailing the proposed feature descriptor, the conventional HOG and two of the most recent improvements of HOG, are discussed as benchmarked methods. For simplicity and speed, linear SVM is used as the baseline classifier throughout the experimentation.

### 2.1. Dataset description/statistics

Many efforts have been expended to standardize the evaluation of performance in machine-vision systems. Although there are available datasets for person detection such as Caltech, INRIA, CAVIAR, etc., there has not been found a dataset of farmers working in fields. Due to this work's intended application, we used UAV to gather working farmers' images in the crop fields. INRIA Persons Dataset and a custom Farmer Dataset (FDS) [30] were used. Originally, INRIA was developed for HOG, which deems it most appropriate to test the proposed method. However, the Farmers' dataset has not been tested before, so to match it with the INRIA dataset, all FDS images are resized to 128 x 64 upright. Besides, multiple scales of images in FDS are also considered. In Table 1, the characteristic of the

datasets used is provided. Only the single-person images from INRIA are used for training and testing, while in FDS, the various poses of the farmers are also included. This allows us to test the robustness of the proposed feature descriptor in discriminating variable poses.

Table 1. Dataset Statistics

Dataset Title	Training		Testing		Total
	Positive	Negative	Positive	Negative	
INRIA	450	1238	568	620	2876
FDS <sup>a</sup>	341	1423	245	1020	3029

<sup>a</sup>. Farmers' dataset is composed of aerial images of farmers working in the field

In Fig. 1, sample images from farmers' Dataset (FDS) are presented. The persons in the upper half are the same in the lower. These images are in very low resolutions and different poses.



Fig. 1. Sample low-resolution images from FDS. Persons in the upper (a) are the same persons in the lower half (b), only they have different poses.

## 2.2. Theoretical comparison of the proposed method to related descriptors: HOG, HEOG, and ImHOG

Histogram of Oriented Gradients is a robust local descriptor designed to capture the gradient information that emphasizes the edges of an object or human figure. It starts with the normalization of gamma and the color of the input image. Then, gradients are computed, which are used for weighted voting into spatial and orientation cells. Contrast is also normalized across the overlapping spatial blocks from the HOG features that are collected and fed to the classifier [20]. Image gradients and its magnitude at any given point are calculated as shown in equations (1) to (4) where  $G_x$  and  $G_y$  are the horizontal and vertical gradients, respectively,  $G_{Mag}$  is the gradient magnitude with  $\theta$  as the orientation.

$$G_x(x, y) = \frac{I(x+1, y) - I(x-1, y)}{2} \quad (1)$$

$$G_y(x, y) = \frac{I(x, y+1) - I(x, y-1)}{2} \quad (2)$$

$$G_{Mag}(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)} \quad (3)$$

$$\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (4)$$

The gradient of the image is divided into cells wherein a single histogram is collected. In the conventional HOG, the cell is usually  $8 \times 8$  pixels and the orientation ranges from 0 to  $\pi$ , which is divided into 9 bins. Cells are grouped into blocks (e.g. four cells per block) and normalized using L2-norm described in (5).

$$L2 - norm, v \rightarrow v / \sqrt{\|v\|_2^2 + \varepsilon^2} \quad (5)$$

Finally, normalized blocks are concatenated in a 1D vector as the HOG feature vector. The conventional HOG is robust in varying illumination; however, its performance is reduced at low-contrast images. HEOG included 3 additional steps on the conventional HOG to address the problem in low contrast. HEOG feature extraction starts with gradient computation on  $x$ - and  $y$ -axes used to derive the gradient's magnitude and orientation. Unlike conventional HOG, the HEOG feature's dimensionality is reduced right after the gradient calculation where only the local maximum on a  $3 \times 3$  running window is considered for gradient voting. If the local maximum is present, the gradient magnitude is preserved, otherwise, the value is zero. Adaptive thresholding is performed by calculating each block's median value to reinstate the robustness to illumination variance lost during the non-max suppression,. The lower (6) and upper (7) thresholds are computed using the median value. These thresholds are used for the next process, i.e. hysteresis thresholding, which preserves the edges affected by illumination. It works by tagging edges as "strong" or "weak". The strong edges participate in the gradient voting, while weak ones are discarded. Lastly, gradient voting and block normalization are conducted, which are similar to conventional HOG's.

$$Th_l = 0.33 * medianvalue \quad (6)$$

$$Th_H = 1.33 * medianvalue \quad (7)$$

Improved HOG addresses the issue of the reversed angles placed in one bin, which is done in conventional HOG rendering some local patterns  $x$  discriminated inadequately. In ImHOG, the disparity between the reverse angles is accumulated across all histogram bins and placed into an extra bin. Initially, Gradient Histogram ( $GH$ ) is generated. Each histogram bin represents a distinct orientation, i.e. reverse angles are placed on distinct bins. Suppose that  $GH(x)$  is the Gradient Histogram value at bin  $x$ , where  $x$  can assume value from 1 to  $N/2$  ( $N$  is the number of bins), and  $GH(x + N/2)$  is the value at the bin opposite to  $x$ . Then, the total disparity over the entire  $GH$  is described in (8).

$$TotalDisparity = \sum_{x=1}^{N/2} \left| GH(x) - GH\left(x + \frac{N}{2}\right) \right| \quad (8)$$

By accounting for the absolute differences (disparities) between bins of reverse angles and converge them into an extra bin, the issue in discriminating similar but high contrast pattern is resolved. However, an extra bin for disparity doesn't work for symmetrical local patterns. That is where the proposed method complements.

In Gabor-enhanced HOG (GeHOG), Gabor filters are used to enhance the HOG at each cell on the ROI. The objective is to resolve the issue of inadequacy in discriminating symmetrical local patterns and those that occur in low contrast. Fig. 2 shows the general framework for generating the Gabor-Enhanced HOG (GeHOG) feature vector.

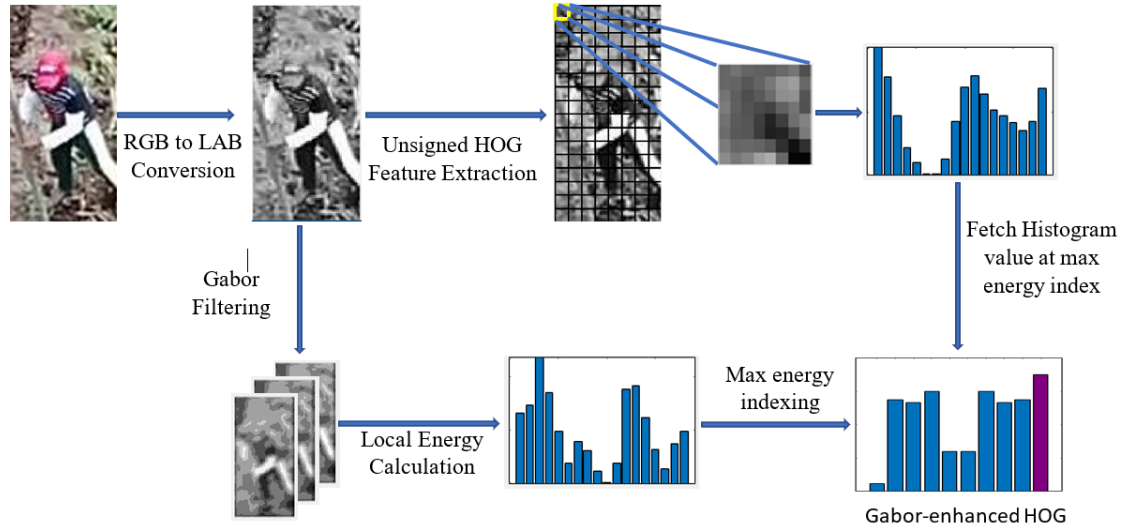


Fig. 2. The proposed framework for extracting feature descriptors.

Initially, a three-channel image is converted to grayscale for reduced dimensionality of the Gabor images which are directly extracted from the single-channel version of the target image. Considering the spatial domain, a 2D Gabor filter resembles a sinusoid-modulated Gaussian kernel described as:

$$G(x, y) = \frac{f^2}{\pi\gamma\eta} e^{\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} + j2\pi f x' + \phi\right)} \quad (9)$$

$$x' = x \cos \theta + y \sin \theta \quad (10)$$

$$y' = -x \sin \theta + y \cos \theta \quad (11)$$

Where  $f$  is the frequency of the sinusoid,  $\theta$  represents orientations orthogonal to the stripes of Gabor function,  $\phi$  is the offset in phase,  $\sigma$  is the standard deviation of the envelope the Gaussian function and  $\gamma$  is the two-dimensional aspect ratio which specifies how elliptical the Gabor function support is.

Gabor filter is applied to each cell in the image's ROI to determine the dominant orientation of edges that occurred. Suppose that there is  $N$  number of Gabor orientations used at  $M$  different scales, then there will be  $N \times M$  Gabor images generated for each cell. For each orientation, cells' local energy at various scales is summed up creating a scalar value denoting the accumulated local energy at that particular orientation. Hence, a 1D vector of length  $N$  is created representing the local energies at different orientations. After normalization, the maximum local energy is used to the index which bin among  $GH$  needs "enhancement". The term "enhancement" refers to increasing the particular orientation's influence in creating the final feature vector. Local energy  $E_s$  at a particular scale  $s$  can be computed using (12) where  $x_{ij}$  is the pixel value at coordinates  $i$  and  $j$  of an  $n \times n$  cell.

$$E_s = \sum_{i=1, j=1}^n |x_{ij}|^2 \quad (12)$$

Local energy histogram is normalized to unity as described in (13) where  $E_{norm}(N)$  and  $E_N$  are the normalized and original value of local energy at  $N^{th}$  bin, respectively, and  $E$  is the 1D vector of local energies.

$$E_{norm}(n) = 0.01 + \frac{E_N - \min(E)}{\max(E) - \min(E)} \quad (13)$$

The maximum value in normalized energy,  $E_{norm}(N)$ , is used to determine which bin in the  $GH$  is enhanced. The algorithm for enhancing the indexed orientation bin is described in Table 2 where  $\beta$ , enhancement modulus, is also introduced.

Table 2. The algorithm on Enhancing Indexed Orientation Bins

Steps	Descriptions
1	<b>Initialize:</b> load calculated Normalized Local Energy, $E_{Norm}$
2	<b>Find:</b> [Value , Index] = $\max\{ E_{Norm} \}$ ; <b>do</b> <b>if</b> Index > $N/2$ ,
	<b>then,</b>
	$\beta = (GH(Index) + E_{norm}(Index))$
3	<b>else,</b>
	$\beta = abs(GH(Index) - E_{norm}(Index))$
	<b>End</b>
	<b>Calculate:</b>
4	$HOG(x) = GH(x) + GH\left(x + \frac{N}{2}\right), \left\{x \mid 1 \leq x \leq \frac{N}{2}\right\}$ ; <b>do</b>
5	add $\beta$ to the elements of HOG, i.e. $HOG(N/2+1) = \beta$ ; <b>do</b>

### 2.3. Experiment set-up

The UAV camera used for data acquisition has a 3840 x 2160 resolution paired with a 3-axis gimbal mechanism for optical image stabilization. The 1280 x 720 resolution is used instead of the maximum to simulate the actual surveillance camera drone's distance to the Point-of-Interest. This approach allows the UAV to fly just 30m above the ground while simulating a 100m flight. The three improvements on HOG, namely HEOG, ImHOG, and GeHOG (proposed) are implemented in a workstation with Intel i7 3.20 GHz processor CPU backed with 8GB RAM. The images acquired from the UAV are forwarded to the ground workstation, where human detection is implemented. The original parameter set-up used in the conventional HOG is also implemented to all feature descriptors, i.e., the cell size of 8 by 8 pixels, block size of 2 by 2 cells, and 9 orientation bins. For the proposed method, the Gabor Filter bank is created using five scales, eighteen orientations that represent the GH bins from  $0^\circ$  to  $360^\circ$  separated by  $20^\circ$  interval, and an eight-pixel wavelength to match the dimension of each cell. Based on the original parameter set-up of HOG, the length of the feature vectors for the improved descriptors are as follows: HEOG has 3,780 values (105 blocks by 4 cells with 9 bins each), while ImHOG and GeHOG have 4200 (105 blocks by 4 cells with 10 bins each). To simply contrast the different feature descriptors, linear SVM is used as the base classifier model. Suppose  $G(x_j, x_k)$  is an element  $(j, k)$  of the Gram matrix, where  $x_j$  and  $x_k$  are  $n$ -dimensional vectors representing observations  $j$  and  $k$  in  $x$ . The linear kernel can be described in (14). Also, five-fold cross-validation is used to all datasets during training to reduce the possibility of overfitting the data.

$$G(x_j, x_k) = x_j' x_k \quad (14)$$

### 2.4. Quantitative evaluation

The F1-score derived from Precision and Recall of the linear SVM classifier models is used to evaluate the feature extraction algorithms' performances. The SVM models are trained separately using HEOG, ImHOG, and GeHOG feature vectors. Precision, recall, and F-score measurements are calculated for each algorithm. Precision is the classifier's ability to detect only the important data points while recall

pertains to the classifier's ability to detect all relevant cases in the Dataset. F1-score is a measure that combines the merit of precision and recall in one number. Precision, Recall, and F1-score can be computed as described in equations (15-17).

$$recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (15)$$

$$precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (16)$$

$$F1-score = 2 \frac{precision * recall}{precision + recall} \quad (17)$$

### 3. Results and Discussion

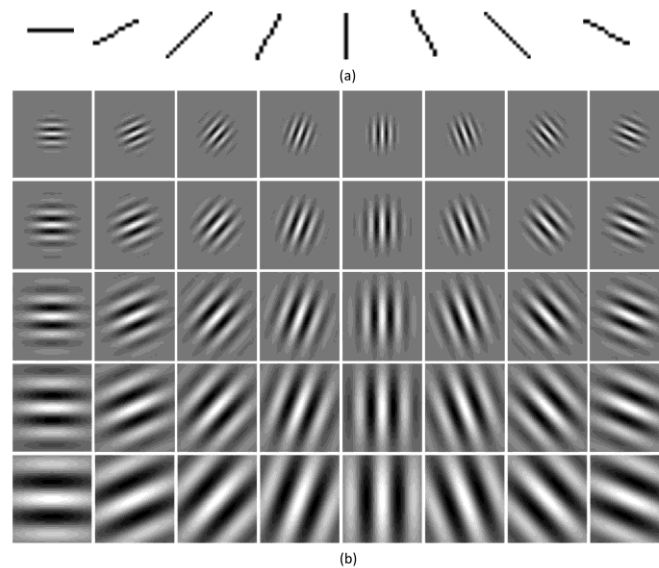
This section presents the results of the experiments done using the proposed and the other two feature descriptors in human detection. Part of the proposed algorithm is the use of enhancement modulus,  $\beta$ , which serves as the discriminating function between GeHOG and other HOG variations. Several functions are evaluated to serve as the enhancement modulus. Table 3 presents the evaluation for the various enhancement moduli derived from Gabor images. The convention used to describe how each modulus is calculated. *The cell* refers to an  $8 \times 8$  region of the Gabor image with *max{}* is finding the maximum pixel value from Gabor cell. The *sum()* refers to the summation of all pixel values while *Mask* is an overlaying kernel and *eigenvalue()* refers to extract the eigenvalue of the  $8 \times 8$  matrix. Fig. 3 shows the orientation masks used as well as the arrays used in generating Gabor cells.

**Table 3.** Percent Precision, Recall and F1-Scores of the classifier using GeHOG with various Enhancement Modulus,  $\beta$

Modulus	Equivalent	Recall	Precision	F1-score
$\beta_1$	$\max\{\text{sum}(\text{Gabor}(\text{cell}))\}$	64.08	98.11	77.53
$\beta_2$	$\text{sum}(\text{Gabor}(\text{cell}))$	74.47	97.69	84.52
$\beta_3$	$\text{sum}(\max\{\text{Gabor}(\text{cell})\})$	64.26	<b>98.12</b>	77.66
$\beta_4$	$\max\{\max\{\text{Gabor}(\text{cell})\}\}$	33.27	94.97	49.28
$\beta_5$	$\text{sum}(\text{Mask} * \text{Gabor}(\text{cell}))$	22.89	95.59	36.93
$\beta_6$	$\text{sum}(\text{eigenvalue}(\text{Gabor}(\text{cell})))$	64.44	96.06	77.13
$\beta_7$	$\max(\text{eigenvalue}(\text{Gabor}(\text{cell})))$	80.46	96.41	87.72
$\beta_8$	$\max\{\text{energy}(\text{Gabor}(\text{cell}))\} - \max\{\text{energy}(\text{Gabor}(\text{cell}) + N/2)\}$	67.96	97.23	80.00
$\beta_9$	$\text{sum}\{\text{energy}(\text{Gabor}(\text{cell}))\}$	78.35	96.53	86.49
$\beta_{10}$	$\text{abs}(\text{GH}(\text{Index}) + E_{\text{norm}}(\text{Index}))$ or $\text{abs}(\text{GH}(\text{Index}) - E_{\text{norm}}(\text{Index}))$	<b>80.99</b>	96.44	<b>88.04</b>

The last row is the modulus selected for enhancing the HOG feature descriptor. It is selected based on the F1-score of the classifier when such is used. Although it renders lower precision than other

moduli, it has the highest recall among the others. The *cell* described in Table 3 has the same size as that of the conventional HOG size. The Gabor cells contain the absolute values derived from filtering.



**Fig. 3.** Orientation masks used to generate Gabor cell with modulus  $\beta_5$  (a), and Gabor arrays used for other moduli (b).

Fig. 4 shows the comparison of the histograms generated using HEOG, ImHOG, and GeHOG. The first column describes the contrasting symmetrical patterns (patterns 1 to 6) in a local  $8 \times 8$  cell. Consider pattern 1 and its invert, pattern 2. The generated HEOG and ImHOG have similar feature histogram generated for the corresponding pair of patterns, manifests the inability of HEOG and ImHOG to discriminate against the contrasting patterns due to the inherent weakness of combining Gradient Histogram's reversed angles. However, this is not the case for GeHOG. It can be noticed that the 10th bin in the histogram changes at the contrast of a pattern. The modulus used to enhance the gradient histogram makes the proposed feature descriptor able to capture the texture and most dominant orientation. With the use of Gabor filtering, GeHOG has enhanced the original HOG feature to discriminate contrasting symmetrical patterns locally adequately. Similar to patterns 3 to 6, the addition of an extra bin in the proposed feature made the HOG patterns distinguishable. The extra bin (10th bin) is influenced by the other bins' local energy and weighted vote as described by the enhancement modulus  $\beta$ . Therefore, the details carried by Gabor images are embedded in that 10th bin. It can also be noticed that the 10th bin in ImHOG cannot discriminate between symmetrical patterns.

Precision, recall, and F1-scores are calculated to evaluate the relevance of the results of the SVM classifier using the feature descriptors HEOG, ImHOG, and GeHOG. In Table 4, the classification F1-score derived from precision and recall is presented. As described, the proposed method performed better than the other two feature descriptors when it comes to the balance of precision and recall.

**Table 4.** Evaluation of the SVM classifier using various feature descriptors (HEOG, ImHOG, and GeHOG) against INRIA and FDS datasets

	INRIA			FDS		
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
HEOG	96.44	80.99	88.04	94.42	75.92	84.16
ImHOG	91.19	83.80	87.34	80.85	46.53	59.07
GeHOG	98.23	78.35	87.17	89.40	79.18	83.98



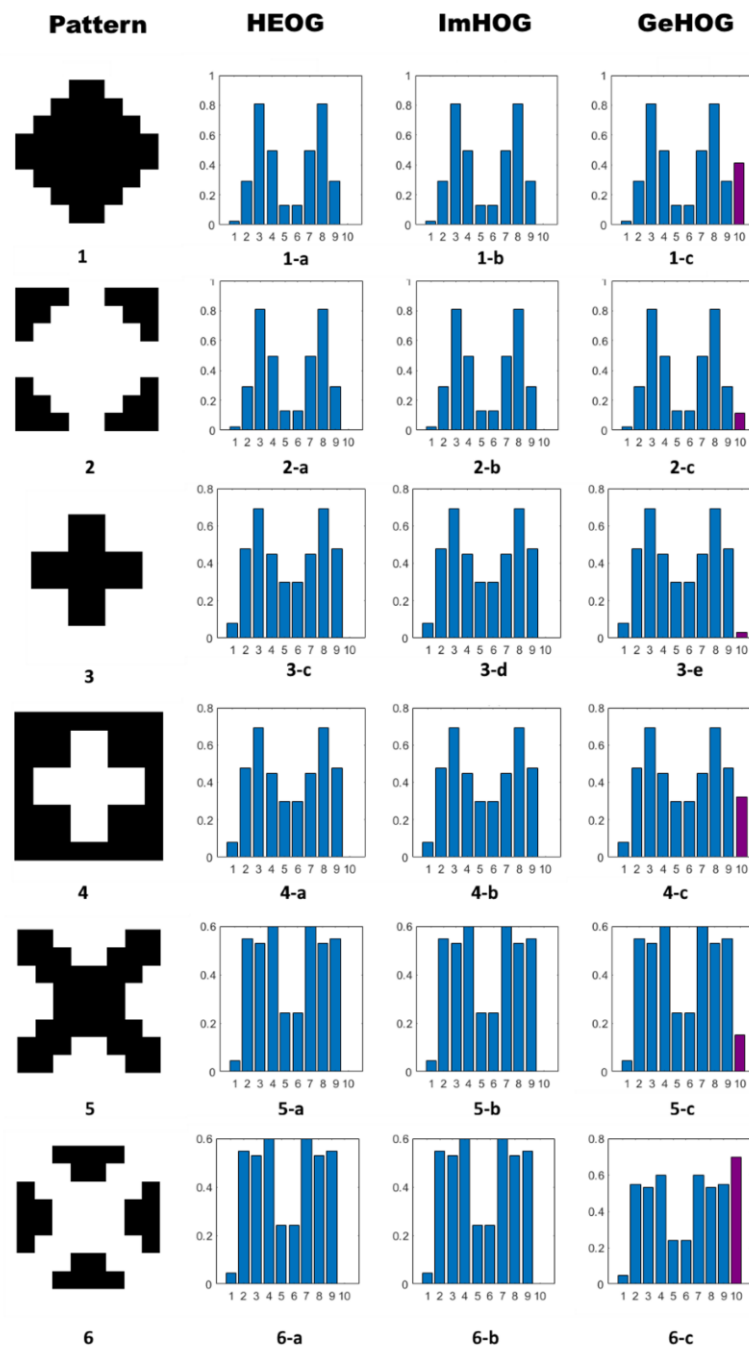


Fig. 4. Comparison of the histograms generated using different feature descriptors. Subplots (1) to (6) are the contrasting patterns and (-a) to (-c) are the corresponding histograms.

#### 4. Conclusion

This paper demonstrates that Gabor filters can be used to enhance a well-known feature such as HOG and its recent improvements. By incorporating the local energy calculations of Gabor images to the HOG feature vector, the influence of such a feature can be increased. This allows any classifier to adequately discriminate the local patterns that are both low in contrast and symmetrical. Initially, Gabor images are generated from a sample image. The number of Gabor images is equal to the scales multiplied by the number of orientation bins. Accumulated local energy per bin is calculated and the bin with the highest energy content is indexed. The indexed bin is a similar bin of Gradient Histogram that is

enhanced. Finally, a new feature vector is generated by adding the enhancement modulus to each cell histogram. GeHOG is used on partial INRIA and FDS. Experiment results show that GeHOG outperformed recent improvements of HOG, i.e. HEOG and ImHOG during the testing process.

However, the diverse pose variations of farmers working in the field remains a challenge and are subject to future works. In aerial images where the camera's position relative to the farmers working is very diverse, the human figure can assume almost all possible poses. The on-going work on this project is developing a monitoring system for farmers' activities based on their poses. The use of GeHOG can provide adequate discrimination for almost similar and symmetrical patterns thus increasing the capability to differentiate poses. Moreover, various dimensionality reduction methods were not implemented in this work but could significantly increase speed. The use of GeHOG in human detection for low-resolution images can be enhanced further by adopting region proposal methods where the trust areas are identified before detection.

### Acknowledgment

The authors thank Gokongwei College of Engineering of De La Salle University (GCOE-DLSU) and Engineering Research and Development for Technology (ERDT) to support this research work.

### Declarations

**Author contribution.** ALPDO performed data acquisition, feature engineering, and analysis. AAB and EPD provided consultations in the conduct of experimentation.

**Funding statement.** The research is funded under COE ERDT Project No. DLSUG 0809.

**Conflict of interest.** The authors declare no conflict of interest.

**Additional information.** No additional information is available for this paper.

### References

- [1] V. Gajjar, Y. Khandhediya, and A. Gurnani, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 2805–2809, doi: [10.1109/ICCVW.2017.330](https://doi.org/10.1109/ICCVW.2017.330).
- [2] J.H. Kim, H.G. Hong, and K.R. Park, "Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light Camera Sensors," *Sensors*, vol. 17, no. 5, p. 1065, May 2017, doi: [10.3390/s17051065](https://doi.org/10.3390/s17051065).
- [3] L. Hou, W. Wan, J.-N. Hwang, R. Muhammad, M. Yang, and K. Han, "Human tracking over camera networks: a review," *EURASIP J. Adv. Signal Process.*, vol. 2017, no. 1, p. 43, Dec. 2017, doi: [10.1186/s13634-017-0482-z](https://doi.org/10.1186/s13634-017-0482-z).
- [4] B. U. Umar, J. Agajo, A. Aliyu, J. G. Kolo, O. S. Owolabi, and O. M. Olaniyi, "Human detection using speeded-up robust features and support vector machine from aerial images," in *2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON)*, 2017, pp. 577–586, doi: [10.1109/NIGERCON.2017.8281928](https://doi.org/10.1109/NIGERCON.2017.8281928).
- [5] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A Review on Human Activity Recognition Using Vision-Based Method," *J. Healthc. Eng.*, vol. 2017, pp. 1–31, 2017, doi: [10.1155/2017/3090343](https://doi.org/10.1155/2017/3090343).
- [6] V. V. Devyatkov, A. N. Alfimtsev, and A. R. Taranyan, "Multicamera Human Re-Identification based on Covariance Descriptor," *Pattern Recognit. Image Anal.*, vol. 28, no. 2, pp. 232–242, Apr. 2018, doi: [10.1134/S1054661818020025](https://doi.org/10.1134/S1054661818020025).
- [7] D. Nguyen, K. Kim, H. Hong, J. Koo, M. Kim, and K. Park, "Gender Recognition from Human-Body Images Using Visible-Light and Thermal Camera Videos Based on a Convolutional Neural Network for Image Feature Extraction," *Sensors*, vol. 17, no. 3, p. 637, Mar. 2017, doi: [10.3390/s17030637](https://doi.org/10.3390/s17030637).
- [8] K. de Miguel, A. Brunete, M. Hernando, and E. Gambao, "Home Camera-Based Fall Detection System for the Elderly," *Sensors*, vol. 17, no. 12, p. 2864, Dec. 2017, doi: [10.3390/s17122864](https://doi.org/10.3390/s17122864).
- [9] M. Buzzelli, A. Albé, and G. Ciocca, "A Vision-Based System for Monitoring Elderly People at Home," *Appl. Sci.*, vol. 10, no. 1, p. 374, Jan. 2020, doi: [10.3390/app10010374](https://doi.org/10.3390/app10010374).

- [10] N. Aldahoul, A. Q. Md Sabri, and A. M. Mansoor, "Real-Time Human Detection for Aerial Captured Video Sequences via Deep Models," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–14, 2018, doi: [10.1155/2018/1639561](https://doi.org/10.1155/2018/1639561).
- [11] R. Llasag, D. Marcillo, C. Grilo, and C. Silva, "Human Detection for Search and Rescue Applications with UAVs and Mixed Reality Interfaces," in *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 2019, pp. 1–6, doi: [10.23919/CISTI.2019.8760811](https://doi.org/10.23919/CISTI.2019.8760811).
- [12] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, "Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN," in *2018 IEEE International Conference on Edge Computing (EDGE)*, 2018, pp. 125–129, doi: [10.1109/EDGE.2018.00025](https://doi.org/10.1109/EDGE.2018.00025).
- [13] K. Liu, W. Wang, and J. Wang, "Pedestrian Detection with Lidar Point Clouds Based on Single Template Matching," *Electronics*, vol. 8, no. 7, p. 780, Jul. 2019, doi: [10.3390/electronics8070780](https://doi.org/10.3390/electronics8070780).
- [14] B. Liu, H. Wu, W. Su, and J. Sun, "Sector-ring HOG for rotation-invariant human detection," *Signal Process. Image Commun.*, vol. 54, pp. 1–10, May 2017, doi: [10.1016/j.image.2017.02.008](https://doi.org/10.1016/j.image.2017.02.008).
- [15] D. Nguyen, H. Hong, K. Kim, and K. Park, "Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras," *Sensors*, vol. 17, no. 3, p. 605, Mar. 2017, doi: [10.3390/s17030605](https://doi.org/10.3390/s17030605).
- [16] G. Masuyama, T. Kawashita, and K. Umeda, "Complementary human detection and multiple feature based tracking using a stereo camera," *ROBOMECH J.*, vol. 4, no. 1, p. 24, Dec. 2017, doi: [10.1186/s40648-017-0092-4](https://doi.org/10.1186/s40648-017-0092-4).
- [17] Y. Yang, Q. Zhang, P. Wang, X. Hu, and N. Wu, "Moving Object Detection for Dynamic Background Scenes Based on Spatiotemporal Model," *Adv. Multimed.*, vol. 2017, pp. 1–9, 2017, doi: [10.1155/2017/5179013](https://doi.org/10.1155/2017/5179013).
- [18] M. Latah, "Human action recognition using support vector machines and 3D convolutional neural networks," *Int. J. Adv. Intell. Informatics*, vol. 3, no. 1, p. 47, Mar. 2017, doi: [10.26555/ijain.v3i1.89](https://doi.org/10.26555/ijain.v3i1.89).
- [19] T. Liu, H. Y. Fu, Q. Wen, D. K. Zhang, and L. F. Li, "Extended faster R-CNN for long distance human detection: Finding pedestrians in UAV images," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, 2018, pp. 1–2, doi: [10.1109/ICCE.2018.8326306](https://doi.org/10.1109/ICCE.2018.8326306).
- [20] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
- [21] P. Hurney, E. Jones, P. Waldron, M. Glavin, and F. Morgan, "Night-time pedestrian classification with histograms of oriented gradients-local binary patterns vectors," *IET Intell. Transp. Syst.*, vol. 9, no. 1, pp. 75–85, Feb. 2015, doi: [10.1049/iet-its.2013.0163](https://doi.org/10.1049/iet-its.2013.0163).
- [22] C.-H. Zheng, W.-J. Pei, Q. Yan, and Y.-W. Chong, "Pedestrian detection based on gradient and texture feature integration," *Neurocomputing*, vol. 228, pp. 71–78, Mar. 2017, doi: [10.1016/j.neucom.2016.09.085](https://doi.org/10.1016/j.neucom.2016.09.085).
- [23] C. Conde, D. Moctezuma, I. Martín De Diego, and E. Cabello, "HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments," *Neurocomputing*, vol. 100, pp. 19–30, Jan. 2013, doi: [10.1016/j.neucom.2011.12.037](https://doi.org/10.1016/j.neucom.2011.12.037).
- [24] Q. LIU, W. ZHANG, H. LI, and K. N. NGAN, "Hybrid human detection and recognition in surveillance," *Neurocomputing*, vol. 194, pp. 10–23, Jun. 2016, doi: [10.1016/j.neucom.2016.02.011](https://doi.org/10.1016/j.neucom.2016.02.011).
- [25] S. Yao, S. Pan, T. Wang, C. Zheng, W. Shen, and Y. Chong, "A new pedestrian detection method based on combined HOG and LSS features," *Neurocomputing*, vol. 151, pp. 1006–1014, Mar. 2015, doi: [10.1016/j.neucom.2014.08.080](https://doi.org/10.1016/j.neucom.2014.08.080).
- [26] D. Sangeetha and P. Deepa, "A low-cost and high-performance architecture for robust human detection using histogram of edge oriented gradients," *Microprocess. Microsyst.*, vol. 53, pp. 106–119, Aug. 2017, doi: [10.1016/j.micpro.2017.07.009](https://doi.org/10.1016/j.micpro.2017.07.009).
- [27] S. A. Chowdhury, M. M. S. Kowsar, and K. Deb, "Human detection utilizing adaptive background mixture models and improved histogram of oriented gradients," *ICT Express*, vol. 4, no. 4, pp. 216–220, Dec. 2018, doi: [10.1016/j.icte.2017.11.016](https://doi.org/10.1016/j.icte.2017.11.016).

- 
- [28] H.-K. Chen, X.-G. Zhao, S.-Y. Sun, and M. Tan, "PLS-CCA heterogeneous features fusion-based low-resolution human detection method for outdoor video surveillance," *Int. J. Autom. Comput.*, vol. 14, no. 2, pp. 136–146, Apr. 2017, doi: [10.1007/s11633-016-1029-8](https://doi.org/10.1007/s11633-016-1029-8).
- [29] A. L. P. de Ocampo, A. A. Bandala, and E. P. Dadios, "Coverage Path Planning on Multi-Depot, Fuel Constraint UAV Missions for Smart Farm Monitoring," in *2018 IEEE Region Ten Symposium (Tensymp)*, 2018, pp. 13–18, doi: [10.1109/TENCONSpring.2018.8691955](https://doi.org/10.1109/TENCONSpring.2018.8691955).
- [30] A. L. P. De Ocampo and E. Dadios, "Radial greed algorithm with rectified chromaticity for anchorless region proposal applied in aerial surveillance," *Int. J. Adv. Intell. Informatics*, vol. 5, no. 3, p. 193, Nov. 2019, doi: [10.26555/ijain.v5i3.426](https://doi.org/10.26555/ijain.v5i3.426).