Available online at www.sciencedirect.com
ScienceDirect
ICT Express 4 (2018) 216–220
www.elsevier.com/locate/icte

Human detection utilizing adaptive background mixture models and improved histogram of oriented gradients

Shayhan Ameen Chowdhury, Mir Md. Saki Kowsar, Kaushik Deb*

Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET), Chittagong-4349, Bangladesh

Received 26 May 2017; received in revised form 7 September 2017; accepted 9 November 2017
Available online 12 December 2017

Abstract

Detecting human is a crux issue in computer vision, with numerous usages especially in human–computer interaction and video surveillance. A framework for human detection with various poses and appearances is proposed in this paper. Initially, a background model is utilized to generate the background image and foreground pixels are classified. Then, HSI color model and color correlogram for removing shadow region and partial occlusion handling are used, respectively. After that, the framework extracts Regions of Interest (ROIs) by analyzing the structure of human body. Finally, features are generated from ROI for classification. A feature descriptor, Improved Histogram of Oriented Gradients (ImHOG), is proposed to alleviate the limitation of HOG. The proposed framework is tested using various videos and the result demonstrates remarkable efficiency with effectiveness.

Keywords: Human detection; Color correlogram; Shadow region; Partial occlusion; Histogram of Oriented Gradients (HOG)

1. Introduction

Human detection is an important aspect of computer vision with vast application areas including event monitoring systems, suspicious event detection, traffic flow measurement, counting human in the crowd, automotive safety and intelligent control. Detecting human in video surveillance system also plays a vital role in fighting crime and protecting public property. Video surveillance is a valuable aid to improve community safety by monitoring important crowded places such as town and city centers, industrial parks, hospitals and universities for early identification of crime and other disruptive incidents. However, with large scale implementation of video surveillance systems manually tracking each camera to identify suspicious events is not possible. Detecting humans in the robust environment is the fundamental process for suspicious event detection. Therefore, we propose a framework for human detection in various poses, appearances, under uneven illuminations and occlusion.

2. Related work

In current human detection frameworks, ROI extraction and feature representation are two main factors being investigated. In [1], intensity difference of individual pixel is incorporated with shape oriented features to capture salient object features. However, the framework selected some important thresholds based on hypothesis. In [2], foregrounds are separated by subtracting background and classified through SVM. However, the framework only detects upper part of human body. In [3], several shape features such as HOG [4], LGP and LBP are combined to a composite local feature. Nonetheless, the extended dimension of the composite feature escalates the processing cost of the system. Other commonly used feature descriptors are EOH, shapelet and Haar wavelet. In case of partial occlusion, it is more efficient to detect human body partly [5] instead of the whole one. However, the improved accuracy also extends the processing cost. Such as the system proposed in [6] roughly

* Corresponding author.
E-mail address: debkaushik99@cuet.ac.bd (K. Deb).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2017.11.016

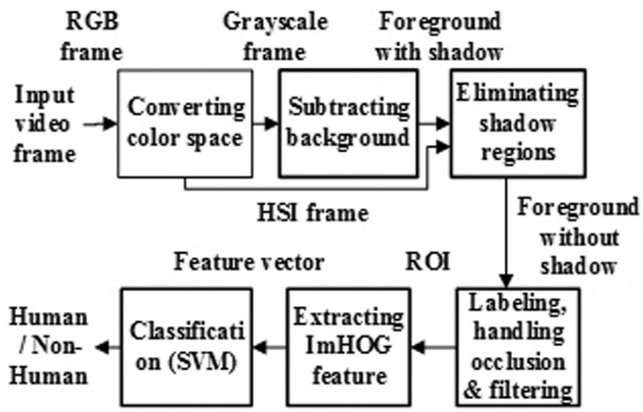

Fig. 1. Proposed framework for detecting human.

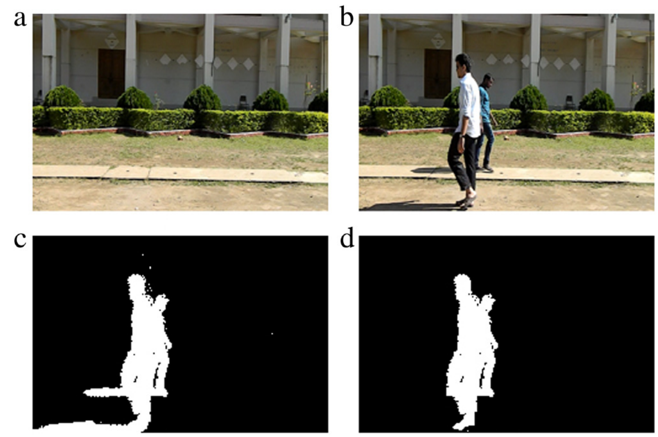

Fig. 2. Shadow elimination instance: (a) background frame, (b) current frame, (c) foreground with shadow image and (d) foreground without shadow image.

needs one minute and seven seconds to operate individual frame.

A framework for detecting human in various environmental conditions is proposed in this paper. The primary emphasis of this paper is to generate a feature vector, which can detect humans in robust environment. The limitation of Histogram of Oriented Gradients (HOG) is studied and identified that it cannot properly differentiate between some different local patterns as shown in Fig. 3, Sample 2(a–c). A feature vector, i.e., improved histogram of oriented gradients (ImHOG) is proposed to alleviate the limitation of HOG by concatenating gradient of opposite directions as presented in Fig. 3, Sample 2(d).

3. Proposed framework

In the author's previous work [7], occlusion handling and human detection based on Histogram of Oriented Gradients (HOG) was presented. This section describes, an extended form of the framework for detecting human by alleviating the limitation of HOG. The proposed framework for detecting human is illustrated in Fig. 1.

3.1. Converting color space

At first, the input frame (RGB) is transformed into grayscale and HSI color space. The grayscale frame is employed to improve the processing speed and the HSI frame is utilized to distinguish shadow pixels from the foreground since HSI color space is more efficient to illumination changes compared to RGB color space.

3.2. Subtracting background

Rather than modeling the value of all pixels by same dispersion, every pixel values are represented as a combination of Gaussians to define various background objects. The framework identifies foregrounds by classifying the Gaussian dispersions which may correlate to background colors. For classification, the framework considers regularity and standard deviation of individual dispersion. In [7], a complete description could be found.

3.3. Eliminating shadow regions

The efficiency of ROI identification depends on distinguishing and removing the shadow regions from the foreground. To detect shadow pixels, a Hue-Intensity disparity (D_{HI}) value is calculated as:

$$D_{HI}(X) = C^{*} H_{Diff}(X) + |\log_e \left(\frac{I_{X,Bg}}{I_{X,Curr}} \right)| \tag{1}$$

where $I_{X,Bg}$ and $I_{X,Curr}$ stand for the intensity of pixel X for background and current frame respectively. C denotes a constant. $H_{Diff}(X)$ represents the absolute difference of hue for pixel X between the current and background frame. Eq. (1), is utilized to identify and eliminate shadow pixels. A more detailed explanation could be found in [7]. The instance of shadow elimination method is described in Fig. 2.

3.4. Labeling, handling occlusion and filtering

The framework labels each binary large object in a group by using color correlogram and back-projection histogram. Then, solidity and aspect ratio of the labeled objects are calculated to eliminate non-human regions.

3.5. Extracting ImHOG feature

The Histogram of Oriented Gradients (HOG) proposed in [4] is a powerful feature vector that uses gradient magnitude and angle information for human detection. HOG is an improvement of the SIFT descriptor proposed in [8] that applied spatial normalization on Gradient Histogram (GH). Dalal and Triggs [4] experimented with both GH and HOG features for object detection and realized that GH discriminates the circumstances where a luminous human region is in front of a dim background and vice versa because the GH deals with gradient directions from 0° to 360°. Fig. 3, shows the situation for the cells of 8 × 8 pixels. For a human detection problem, this discrimination causes a huge intra-class variation. Dalal and Triggs [4] resolved the issue related to GH by calculating

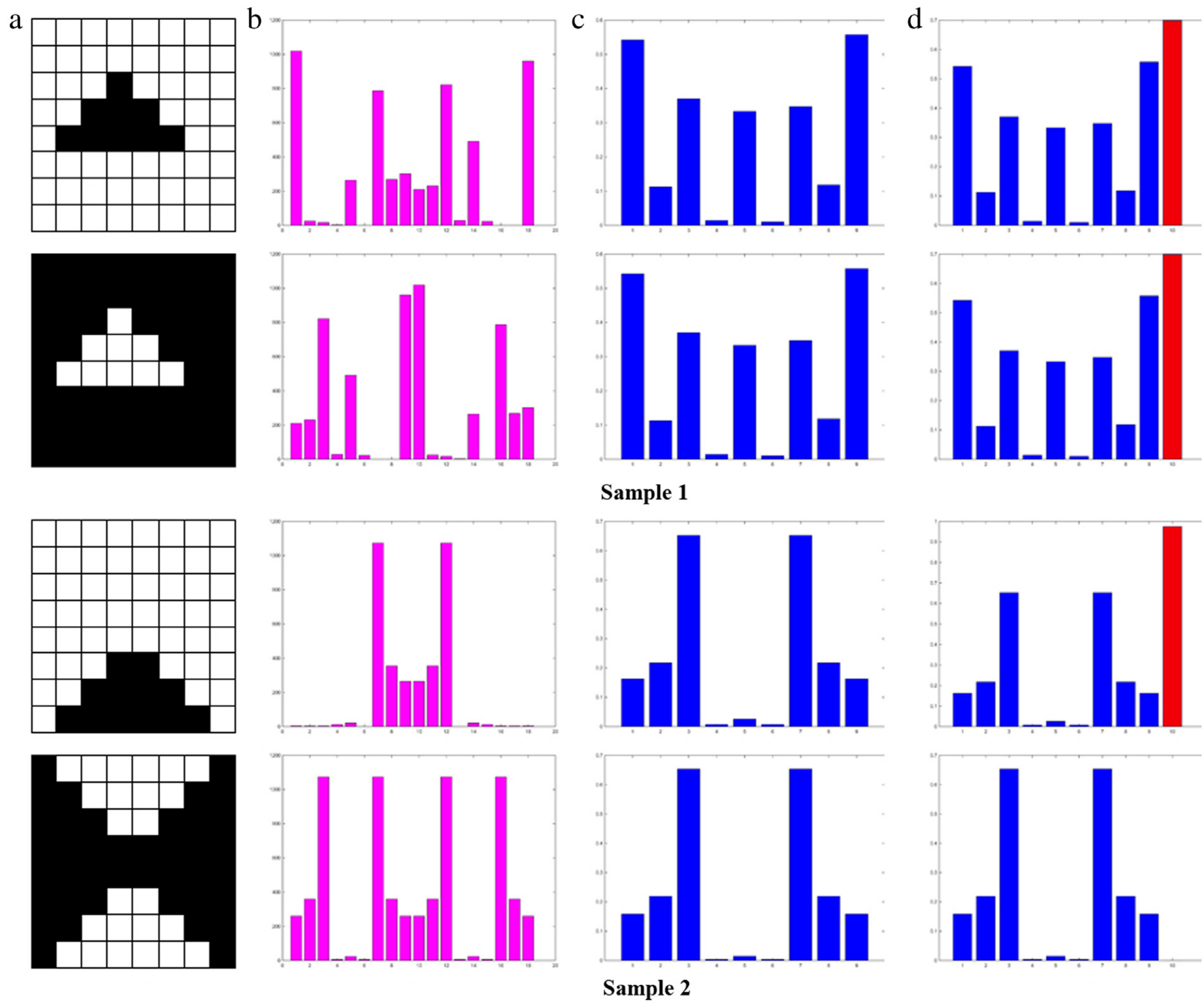

Fig. 3. Representation of different local patterns by various feature vectors: (a) Pattern, (b) GH, (c) HOG and (d) ImHOG.

gradients of angle α and $\alpha + 180°$ (reversed orientation) to α only, where $0° \leq \alpha < 180°$. Patterns in Fig. 3, Sample 1 are represented differently by GH. However, HOG represents those patterns identically as illustrated in Fig. 3, Sample 1(c). As a result, HOG can detect human region regardless of a bright or dark background. As HOG puts angles of reversed orientations to one histogram bin, some local patterns cannot be discriminated adequately by HOG. Thus, it is possible for two distinct patterns to be represented by an identical HOG feature vector. Let $GH(x)$ denotes the value of bin x for sampled gradient angle α and M represents the number of bins in GH. HOG is the sum of two corresponding bins of GH. Therefore, HOG can be computed from GH by adding $GH(x)$ and $GH(x + M/2)$, where $1 \leq x \leq M/2$. As HOG minimize GH, some key features are lost. To resolve the previously stated issue related to HOG, a new histogram GH_{Dis}, called histogram of gradient disparity is generated by taking the absolute difference between $GH(x)$ and $GH(x + M/2)$, where $1 \leq x \leq M/2$. Then the values of all the bins of GH_{Dis} are summarized into one bin called $exBin$. This new bin $exBin$ can adequately discriminate the local patterns which are misclassified by HOG. After that, HOG and $exBin$ are concatenated for every cell in the ROI to generate improved histogram of oriented gradients (ImHOG). ImHOG can correctly discriminate local patterns misclassified by HOG. Fig. 3, Sample 2 explains the situation where HOG represents two distinct patterns by an identical feature vector. In Fig. 3, both patterns of Sample 2(a) are represented similarly by HOG as illustrated in Fig. 3, Sample 2(c). However, the $exBin$ included in ImHOG provides different values for the patterns similarly represented by HOG as shown in Fig. 3, Sample 2(d). As a result, ImHOG can differentiate those patterns. Furthermore, ImHOG also resolves the issue of GH where a luminous human region is in front of a dim background and vice versa as shown in Fig. 3, Sample 1. In Fig. 3, for both patterns of Sample 1(a) ImHOG generates the same feature as shown in Sample 1(d). Thus ImHOG not only properly represent the patterns misclassified by HOG, also ImHOG does not discriminate the circumstances where a luminous human region is in front of a dim background and vice versa.

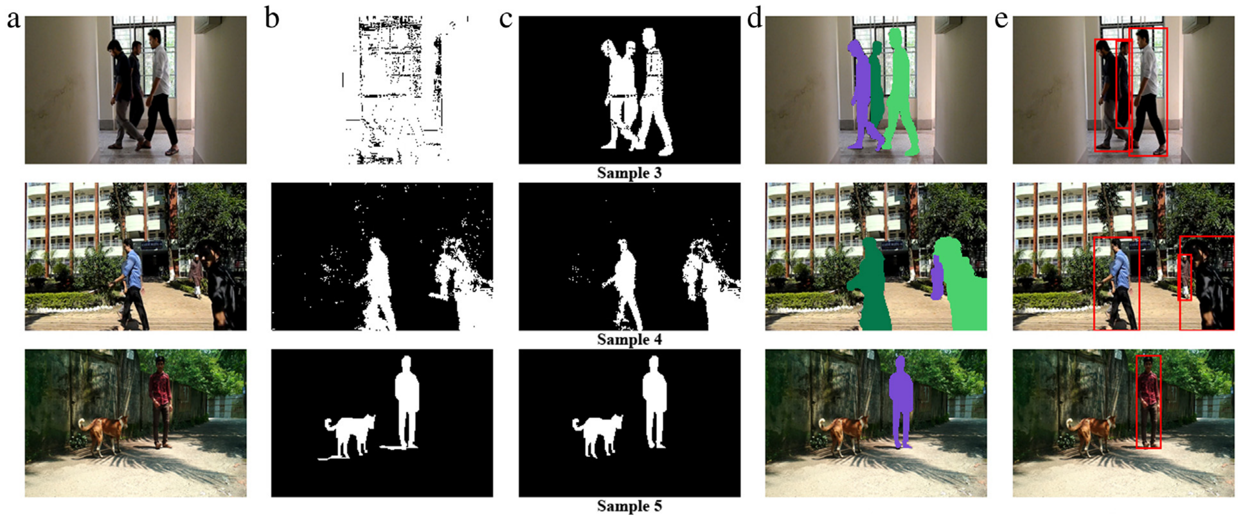

Fig. 4. Processing example for detecting human: (a) RGB frame, (b) foregrounds with shadows, (c) foregrounds without shadows (d) labeled foreground(s) and (e) detected object(s).

Table 1
Precision and recall values with different environmental conditions.

Frame type	Total frame	Pre (%)	Rec (%)	Avg time (s)
Indoor	489	94.8	93.7	0.46
Outdoor	675	93.5	93.1	
Complex background	262	92.7	89.6	
Average	1430	**93.7**	**92.1**	

3.6. Classification

The ImHOG feature vector is passed through a linear SVM for detecting human, where SVM is a supervised margin classifier. For two clustered training dataset, SVM aims to discover maximum margin hyperplane, that process biggest segregation between the clusters.

4. Experimental results

This section explains the experimental outcomes of the proposed framework for detecting human. Experiments are accomplished on Intel Core i5 3.20 GHz CPU with 4 GB RAM using MATLAB environment. Input videos have been captured with a static camera (Nikon L120) at a rate of 25 fps and a resolution of 320 × 240 pixels. These videos contain humans in various environmental conditions. 140 and 1430 frames are utilized to train and test the proposed framework, respectively. Nonetheless, these test sequences are different from the training sets.

Fig. 4 shows the processing example of the proposed framework. In that figure, Sample 3 contains occlusion with three humans in the indoor environment. As light comes from behind the humans, shadows of the humans are projected on the wall. As a result, the background subtraction process cannot detect the foreground region properly as shown in Fig. 4, Sample 3(b).

However, the proposed shadow elimination process detects and removes the shadow regions from foreground as shown in Fig. 4, Sample 3(c). Fig. 4, Sample 4(e) reveals that the framework can detect humans in different appearances. Finally, in Fig. 4, Sample 5(d) the dog is discarded from ROI as the shape of the dog does not match the filtering conditions.

Table 2
Result comparison.

Framework	TP	FP	FN	Pre (%)	Rec (%)
Proposed framework	1343	90	115	93.7	92.1
[2]	891	156	367	85.1	70.8
[4]	1156	107	125	91.5	90.2

The precision and recall value is computed from different types of video frames. Table 1, illustrations the precision (Pre) and recall (Rec) values in different environmental circumstances. The proposed framework demonstrates a superior result to indoor and outdoor video frames and also exhibits acceptable response for video frames covering complex background. Results generated from the proposed framework are compared with [2] and [4]. The comparison is accomplished on precision and recall value in Table 2. In Table 2, FP, FN and TP denote False Positive, False Negative and True Positive respectively. Results show in Table 2 that our framework explicitly improves the performance with respect to proposed framework in [2]. The framework presented in [2] is not robust enough to detect human in various poses, appearances, under uneven illuminations and occlusion. Furthermore, the proposed framework provides marginally better detection result than the method proposed in [4]. The recall values of Table 2 justify that the proposed framework outperforms for detecting human even in robust environment.

The performance of the proposed feature vector ImHOG is compared with [4] using INRIA human dataset. Fig. 5, post the

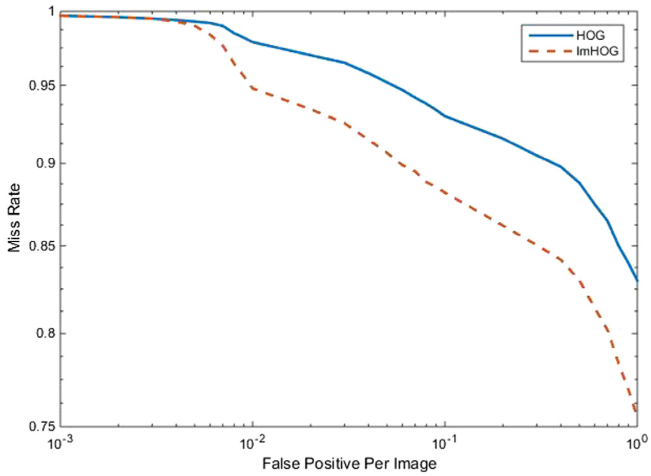

Fig. 5. Performance comparison between ImHOG and HOG.

miss rate against the false positive per image (FPPI) for both features. The lower the curve, the better the performance. At 10^{-1} FPPI, ImHOG rank first followed by HOG. From Fig. 5, it can be seen that ImHOG consistently outperforms over HOG.

5. Conclusion

A framework for human detection is proposed in this paper by generating a feature vector i.e. ImHOG which can differentiate local patterns properly. ImHOG is derived to alleviate limitations of GH and HOG. GH differentiates a dark ROI in front of a bright background and vice versa. HOG generates the same feature for some different local patterns. The proposed feature vector ImHOG solves these problems by adding a new bin, i.e., exBin to HOG. The exBin could distinguish patterns misclassified by HOG and does not separate a dark ROI in front of a bright background and vice versa. As a result, the proposed feature vector ImHOG alleviates the limitations of GH and HOG. However, the proposed framework assumed backgrounds to be entirely static (like rain, a waving tree branch confuses the proposed framework). The changing background problem can be solved through analysis of temporal variations.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Y. Ma, L. Deng, X. Chen, N. Guo, Integrating orientation cue with eoh-olbp-based multilevel features for human detection, IEEE Trans. Circuits Syst. Video Technol. 23 (10) (2013) 1755–1766.
[2] R. Tong, D. Xie, M. Tang, Upper body human detection and segmentation in low contrast video, IEEE Trans. Circuits Syst. Video Technol. 23 (9) (2013) 1502–1509.
[3] B. Jun, I. Choi, D. Kim, Local transform features and hybridization for accurate face and human detection, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1423–1436.
[4] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proc. IEEE Int. Conf. CVPR, June 2005, vol. 1, pp. 886–893.
[5] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
[6] A. Vedaldi, V. Gulshan, M. Varma, A. Zisserman, Multiple kernels for object detection, in: Proc. IEEE Int. Conf. CVPR, 2009, pp. 606–613.
[7] S.A. Chowdhury, M.N. Uddin, M.M.S. Kowsar, K. Deb, Occlusion handling and human detection based on histogram of oriented gradients for automatic video surveillance, in: Proc. International Conference on Innovations in Science, Engineering and Technology (ICISET), 2016, pp. 1–4.
[8] D.G. Lowe, Distinctive image features from scaleinvariant keypoints, Int. J. CV 60 (2) (2004) 91–110.