2nd International Conference on Electrical, Computer and Communication Engineering (ECCE),
7-9 February, 2019

Human Detection Based on HOG-LBP for Monitoring Sterile Zone

Farjana Bintay Kamal[1*], Sangita Mitra[2], Tahmina Khanam[3], Uzzal K. Acharjee[4], Dipankar Das[5], Kaushik Deb[6]
[1, 2, 3, 6] Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET)
Chattogram, Bangladesh
[E-mail: {farjanarema13[1], sangitamitra75cuet[2]}@gmail.com, {tahmina_iict[3], debkaushik99[6]}@cuet.ac.bd]
[4] Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh
[E-mail: uzzal[4]@cse.jnu.ac.bd]
[5] Department of Information and Communication Engineering, University of Rajshahi, Rajshahi, Bangladesh
[E-mail: dipankar[5]@ru.ac.bd]

***Abstract*— Human detection has enormous application area in autonomous video surveillance and human-computer interaction. Detecting suspicious event has become a very crucial issue in the current circumstance of our society. As a pioneer, a framework is proposed for detecting human in the sterile zone in this paper. Since, in the case of the sterile zone, we have to deal with low-resolution video, that's why initially input video frames are enhanced by using local histogram equalization. Then a background model is created by using the Gaussian Mixture Model (GMM) where each pixel is represented by a mixture of a number of Gaussian based on probabilistic method. This modeled background is then compared with a new frame to detect the foreground object. After that, the morphological operation is performed to remove discontinuities and to get the region of interest (ROI). Then shape and texture features from ROI are extracted for classification. Finally, combined features from Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP) are fed into SVM classifier to detect human. In this paper to achieve better performance in the sterile zone, human shape is analyzed with HOG along with enumerating local features by LBP. Moreover, this proposed framework is tested using various video in different conditions and the outcome demonstrates remarkable efficiency comparative to other alternatives.**

Keywords*— Sterile zone monitoring, Gaussian Mixture Model, HOG, LBP, SVM.

I. INTRODUCTION

Over the recent years, detecting human in video surveillance system has become a demanding research field of computer vision. Monitoring sterile zone for detecting suspicious event is the most challenging issue since surveillance videos are usually low resolution video. Generally, uneven illumination condition and cluttered background exacerbate this situation. Human handled surveillance system depends on human to monitor the screen for detecting suspicious incident in the restricted area which requires infinite time. However, there can be happened any unpleasant event because there are limitations in human capability. The enormous application area of sterile zone monitoring includes border surveillance, prison fence monitoring, ATM booth monitoring and ensuring safety in any restricted area.

In the past few years, application of human detection for video surveillance has been actively developed. Human detection is difficult because of different shape, appearance and color of human. For human detecting foreground and extracting feature are two important factors. For monitoring sterile zone a method is presented in [1] in which Gaussian mixture model is used for background modeling. Although, this procedure can identify a foreground, this is not so convenient in case of low quality video because no preprocessing technique has applied to enhance images. Background subtraction to find the foreground object and feature extraction are two most challenging work in the field of human detection. To detect unattended object a method based on dual background difference is proposed in [2] which depends on reference background model. But this gives false results in case of abrupt change of lighting and weather conditions. In [3], a method is presented where background model is created by cumulative dual foreground difference for video surveillance. Not only background model but also temporal analysis, vehicle detector and tracking are also working in this video surveillance. However, this method is not adaptive with the noisy foreground images and, as a result it produces many false alarms. The problem of detecting stationary object is solved in [4] by testing the stability of pixels and after that, extracting stable regions from the image. But the main drawback of this procedure is it generates false alarms when data from background subtraction are calculated inaccurately. For road traffic surveillance system a method is proposed in [5] based on three dimensional information of suspected abandoned object. Though this new algorithm has improved performance to a great scale, it is very time consuming. In [6] HOG feature is used to get the shape appearance of human. As we know that combined features are always better than single feature. Since HOG only gives information about the shape of human, by using LBP we get the texture of human.

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

Therefore, we propose a framework for human detection to monitor sterile zone from low resolution video considering uneven illumination and occlusion. After detecting foreground from enhanced image by using Gaussian mixture model, we have calculated feature vector by using both HOG and uniform LBP to achieve better performance. After that, combined feature vectors are fed into linear support vector machine to classify that whether the detected object is human or not.

Rest of the paper is summarized as follows: The proposed framework for human detection is described in section II. Simulation results are represented in section III and section IV encloses the concluding remarks and future works.

II. PROPOSED FRAMEWORK FOR HUMAN DETECTION

In this section the proposed framework of human detection has been explained in details. The proposed framework consists of eight steps: 1.Converting RGB frames to Gray, 2.Enhancing using local histogram equalization, 3.Detecting foreground, 4.Labeling and filtering, 5.Extracting HOG feature, 6.Applying uniform LBP, 7.Combining HOG and LBP features and 8.Classification using SVM. Figure 1 shows the proposed framework of human detection for monitoring sterile zone.

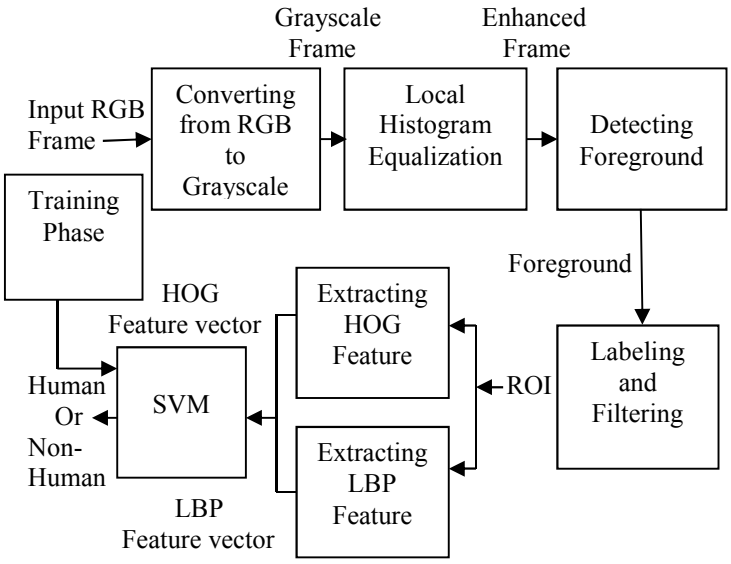


Fig. 1. The proposed framework for detecting human.

A. Converting from RGB to Grayscale frame

RGB frames are extracted from input video and transformed into grayscale frames. The number of pixels reduced from approximately 2^24 to 256 as a result the processing speed is reduced.

B. Enhancing using local histogram equalization

Basically histogram equalization spreads out intensity values along the total range of values to achieve higher contrast

(a) (b)

Fig. 2. Processing examples of local histogram equalization: (a) input RGB frame, and (b) enhanced frame.

[7]. That's why it enhances the whole image unnecessarily, moreover, it arises a chance to miss small details over the image. In contrast, Local histogram equalization spreads out most frequent intensity values in an image and enhances details over small area in an image [8]. As for monitoring sterile zone, we have to process low resolution video, it is necessary to enhance image contrast by applying local histogram equalization. The processing example of local histogram equalization is shown in figure 2.

At first, the gray scale image is padded with zero on all sides and a $n \times n$ window size is considered and it passes over the image to enhance the low contrast image.

The window starts from the position (1,1) and the size of window can be changed as required. For the particular window we have to find probability density function (PDF) and cumulative distribution function (CDF) for each pixel. Then multiply all these CDF by highest pixel vale of entire image and after rounding these values we get the enhanced value of middle pixel of the particular window. Following same procedure for each window we get the enhanced frame.

C. Detecting foreground

After getting the enhanced frame, background subtraction is performed to get the detected foreground. Usually, all pixel values are modeled with a particular distribution. However, we have modeled a specific pixel with a mixture of Gaussians. Generally, 3 to 5 Gaussians are used to model the background. To classify a particular pixel between background and foreground, the value of mean and variance of that pixel is calculated. If any pixel doesn't fit with the background distribution then it is classified as foreground pixel. The probability of occurrence of a particular pixel at time t is expressed as 1.

$$p(A_t) = \sum_{i=1}^{K} w_i . \eta(A_t, \mu_i, \sigma_i^2) \quad (1)$$

A_t is the observed pixel at i-th frame at time t and w_i, μ_i and σ_i^2 are weight, mean and standard deviation of that pixel. The probability density function η is defined as (2).

$$\eta(A_t, \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\Pi}\sigma} e^{-\frac{(A_t-\mu)^2}{2\sigma^2}} \quad (2)$$

(a) (b)

Fig. 3. Processing examples of background subtraction: (a) input enhanced frame, and (b) background subtracted frame.

After getting the modeled background, it is subtracted from new frame to detect foreground object. If the pixel's standard deviation is greater than 2.5 from the mean value of background model, then the pixel value is denoted as foreground otherwise it will considered as background. The processing example of background subtraction is illustrated in figure 3. After detecting each foreground, we need to update the background model. In [1] a complete description of update could be found. This procedure of detecting foreground and updating background will be repeated simultaneously till the end of sequence frame.

D. Labeling and filtering

Before labeling and filtering, morphological operations are performed to remove undesirable noise and holes of foreground frame. Opening reduces undesired noise while at the same time deforms the shape of the object. That's why closing is performed which fills holes and produces clean foreground which is sent to labeling and filtering step. The process of labeling and filtering is done by labeling the entire region. Then the regions which are small are removed by filtering and get the region of interest. The process of labeling and filtering begins by inverting the binary image in order to speed up the processing time. Then, concentrate boundaries are identified. After that shape properties are determined from the ratio of dimensions and roundness. Finally shapes are classified according to its properties. Figure 4 shows the processing example of labeling and filtering.

E. Extracting HOG features

In our experiment, features of region of interest are extracted by Histogram of Oriented Gradient (HOG) [6]. For HOG feature, RGB image is needed. For that reason we have superimposed the clean foreground and RGB image. From this, RGB image of foreground is found. At first, the image is resized into 128*64 pixels.

(a) (b)

Fig. 4. Processing examples of labeling and filtering: (a) input foreground frame, and (b) labeled foreground frame.

Gradient magnitude, $M = \sqrt{S_x^2 + S_y^2}$ (3)

Gradient orientation, $\Theta = \arctan\left(\frac{S_y}{S_x}\right)$ (4)

Then the gradient of the superimposed image is computed with a 1-D kernel both in the horizontal and vertical direction. These kernels are denoted by [-1 0 1] & [-1 0 1] ^T respectively. The value of gradient magnitude and orientation is calculated as 3 and 4 respectively. Here, S_x and S_y are gradient value in x direction and y direction respectively.

The sliding window is divided into cells, which is the size of 8*8 pixels. From each cell, a histogram is computed by voting the gradient magnitude of each pixel into 9 bins where each bin corresponds to 20 degrees. Four cells are grouped into blocks where 50% blocks are overlapping. After the concatenation of the histograms of four blocks the vector is normalized to make it contrast invariant. Finally, the block histograms are concatenated into 3780 feature vector in total.

F. Extracting uniform LBP features vector

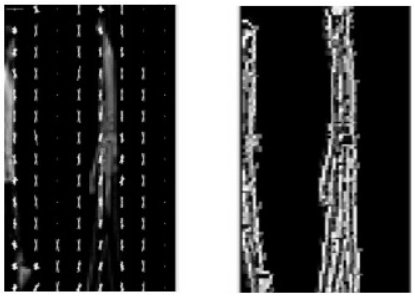
HOG features descriptor gives information about local edge shape of human. However, to detect human more accurately and efficiently, the texture information of human body is also needed. In our proposed framework uniform LBP [9] is used as the texture descriptors. In LBP, an image is transformed into an image of integer labels which describes textures of the image. A pattern and its strength are assumed as two aspect of texture whose are complementary. In the beginning, center pixel of the block thresholds its neighbor pixels value. For the conversion from binary to the decimal of center pixel, the computed value is multiplied by powers of two. After that the values are summed to obtain a label for the center pixel. By following this procedure, 256 different labels are obtained. The number of uniform model is 58, and other model as a class, we can get 59-D vector. The values are calculated as (5).

$$LBP_{P,R} = \sum_{P=0}^{P-1} s(g_p - g_c).2^p \quad (5)$$

Where, g_p is neighborhood pixels in each block, g_c is center pixel value, p is sampling points and R is radius.
If the center pixel's value is greater than the neighbor's value, then the binary threshold function t(x) of (6) will be 0. Otherwise, it will be 1. In this way, 8 digit binary number will be generated.

$$t(x) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases} \quad (6)$$

If the value of U in (7) is less than equal 2 then that pattern is labeled as uniform pattern otherwise non-uniform pattern. LBP features of the four cells in each block are concatenated into a 59-D feature vector of the block. Then after combining with HOG feature vector they are saved in matrix format where each of the images has 3839 features vector. The processing example of HOG and LBP combined features are represented in fig. 5.



(a) (b)

Fig. 5. Processing examples of extracting HOG and LBP feature: (a) HOG feature vector, and (b) LBP feature vectors.

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_p - g_c)| + \sum_{p=0}^{p-1} s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (7)$$

G. Classifying with Support Vector Machine (SVM)

Finally, the combined feature vector from HOG and LBP is passed to Linear SVM for detecting whether the region of interest is human or non-human. SVM is a supervised learning model. For linearly separable training data SVM aim to find hyper-planes that are parallel which separates the data into two classes. Distance between the hyper-planes is as large as possible.

III. EXPERIMENTAL RESULTS

A. Dataset and Experimental Environment

This section illustrates the experimental outcomes of the proposed framework for detecting human. Experiments are accomplished on Intel core i3 2.00 GHz CPU and 4.00 GB RAM using MATLAB environment. We have used a static camera (Canon EOS 1300D) to capture video at a rate of 30 fps and resolution of frames are 320 * 240 pixels. Videos are captured considering different environmental conditions. Our system is trained with 356 frames and tested on 929 frames. Among them some positive and negative training samples are shown in figure 6. Testing frames are chosen considering different illumination conditions and occlusions. Moreover, the training sequences are different from test sets. Further, we have tested our system on our own dataset and i-LIDS dataset.

B. Performance Evaluation Matrix

For performance evaluation, the value of precision and recall is computed from different types of video which are captured by considering different environmental condition and illumination changes. Precision is the ratio of relevant observations to the retrieved instances whereas recall gives the observation to the retrieved instances whereas recall gives the measurement of how many truly relevant results are retrieved. In case of precision and recall, TP, FP, TN and FN indicates the value of true positive, false positive, true negative and false negative respectively.

Fig. 6. Training samples to train SVM with positive samples, and negative samples.

When ROI is correctly detected as human, it is TP and if ROI in incorrectly leveled as human then it is FP. However, if ROI is incorrectly leveled as not human then it is FN and when ROI is correctly detected as not human then it is TN.

To calculate the accuracy of our proposed framework we have also computed F_1 score which is the weighted average of precision and recall. F_1 score is a useful parameter to measure the accuracy in case of uneven class distribution. The value of precision, recall and F_1 score are calculated as follows.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \times P \times R}{(P + R)}$$

C. Experimental Analysis

Our proposed framework is performed in uneven illumination condition and low contrast video for both simple and complex background frames. Here, processing examples from different input video and performance analysis with another framework have discussed. Figure 7 illustrates processing examples of proposed framework to detect human in sterile zone from four different videos. From 1st, 2nd and 3rd video in figure 7, we have detected single multiple human successfully. However, in 2nd video because of occlusion, multiple human wouldn't segmented accurately.

(a) (b) (c) (d) (e)

Fig. 7. Processing example of human detection in sterile zone: (a) input RGB frame (b) enhanced image (c) foreground (d) labeled foreground, and (e) detection.

(a) (b) (c)

(d) (e) (f)

Fig. 8. Sequence of detected frames.

The values of precision and recall values supports this because the F_1 score gives 91.02% accuracy in our framework while the proposed framework in [10] gives 77.64% accurate results on own and i-LIDS dataset.

In figure 8, humans are detected from consecutive frames of testing dataset in sterile zone. Here in a, b and c are multiple human detected from a given video.

Table 2, shows a comparison between our proposed framework and [10] in terms of precision and recall. Furthermore, the proposed framework provides marginally better detection result than [10] as it fails to detect human in under uneven illuminations condition.

TABLE I. PERFORMANCE ANALYSIS ATDIFFERENT ENVIRONMENTAL CONDITION

Frame type		Environment conditions	Total frame	Precision (%)	Recall (%)
Simple background	Single Human	Normal condition	256	92.10	89.8
	Multiple human	Normal condition, low contrast, uneven illumination	168	91.46	90.6
Complex background	Single human	Normal condition, low contrast, uneven illumination	298	92.21	90.8
	Multiple human	Normal condition	207	91.35	89.6

Table 1, summarizes the performance analysis of our proposed framework under simple and complex background along with single and multiple human. For complex background-single human frame type comparatively better precision and recall has been achieved. Average of total frame, precision, and recall for all types of frames are also calculated. When frame type is the complex background

multiple human result for precision and recall is not up to the mark though we have considered only normal condition.

TABLE II. COMPARITIVE RESULTS

Framework	TP	FP	FN	Precision (%)	Recall (%)	F_1 Score
Proposed	760	68	82	91.79	90.26	91.02
[10]	754	116	312	86.7	70.3	77.64

In addition, Receiver Operating Characteristic (ROC) curve of proposed framework is measured for performance evaluation in shown in figure 9, where x-axis and y axis represents false positive rate (FPR) and true positive rate (TPR) respectively. We have used our own dataset and i-LIDS dataset to analysis performance. The value of TPR and FPR are calculated as follows.

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

The value of TPR and FPR are calculated for both proposed framework and [10] in table III. From this data ROC curve is plotted in figure 9. From this ROC curve we can conclude that our proposed framework is performing better than [10].

TABLE III.RECEIVER OPERATING CHARACTERISTIC

Framework	TPR	FPR
Proposed	90.26	78.16
[10]	70.73	54. 205

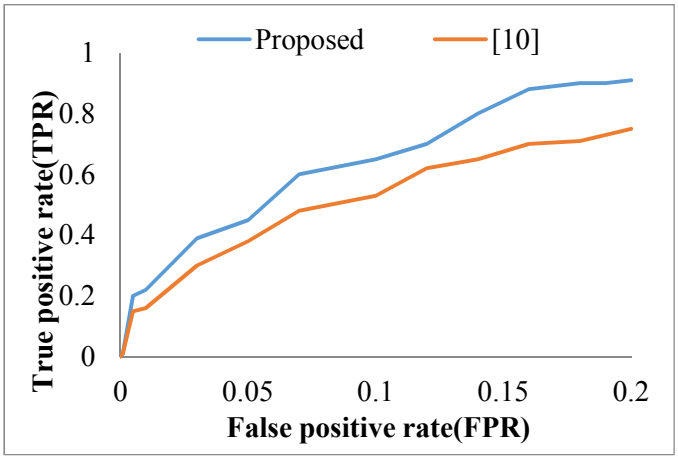


Fig. 9. ROC curve of proposed method and [10].

IV CONCLUSION

A framework is proposed in this paper for detecting human in case of monitoring sterile zone with higher adaptability. Initially, input RGB video frames are converted into grayscale frames and local histogram equalization has performed to enhance the contrast of the frame. It gives us more chance to detect human from low contrast video considering illumination conditions. After that, background model is created and background subtraction has performed to detect foreground object and after labeling and filtering, intended foreground notation is found. We combine contour information and texture information and propose the cascade of the two types of features, HOG features and LBP feature as the feature set because the single feature has a problem of the low recognition rates. Finally, HOG and uniform LBP combined features are sent to linear SVM to classify whether detected region is human or not. The proposed framework is limited to detect humans from videos provided by a static camera. The proposed framework is sensitive to detect multiple human when any occlusion occurred. In future works, this framework could be extended to detect humans from dynamic background and occlusion scenario.

REFERENCES

[1] Shahbaz, Ajmal, and Kang-Hyun Jo. "Sterile zone monitoring with human verification." In Human System Interactions (HSI), 2017 10th International Conference on, pp. 60-63. IEEE, 2017.

[2] Filonenko, Alexander, and Kang-Hyun Jo, "Unattended object identification for intelligent surveillance systems using sequence of dual background difference," in IEEE Transactions on Industrial Informatics 12, no. 6 , 2247-2255, 2016.

[3] Jo, Kang-Hyun. "Cumulative dual foreground differences for illegally parked vehicles detection," IEEE Transactions on Industrial Informatics 13, no. 5, 2464-2473, 2017.

[4] Szwoch, Grzegorz. "Extraction of stable foreground image regions for unattended luggage detection," Multimedia Tools and Application 75, no. 2, 761-786 , 2016.

[5] Zeng, Yiliang, Jinhui Lan, Bin Ran, Jing Gao, and Jinlin Zou. "A novel abandoned object detection system based on three-dimensional image information," Sensors 15, no. 3 (2015): 6885-6904.

[6] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005.

[7] Yang, Jinwen, Weihe Zhong, and Zheng Miao. "On the Image enhancement histogram processing." In Informative and Cybernetics for Computational Social Systems (ICCSS), 2016 3rd International Conference on, pp. 252-255. IEEE, 2016.

[8] Kaur, Hardeep, and Jyoti Rani. "MRI brain image enhancement using Histogram equalization Techniques." In Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, pp. 770-773. IEEE, 2016.

[9] Pan, Zhibin, Hongcheng Fan, and Li Zhang. "Texture classification using local pattern based on vector quantization." IEEE Transactions on Image Processing 24, no. 12 (2015): 5379-5388.

[10] Tong, Ruofeng, Di Xie, and Min Tang. "Upper body human detection and segmentation in low contrast video." IEEE Transactions on Circuits and Systems for Video Technology.