

# Baggage Recognition in Occluded Environment using Boosting Technique

**Tahmina Khanam<sup>1</sup> and Kaushik Deb<sup>1\*</sup>**

<sup>1</sup>Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET)  
Chittagong, 4349 - Bangladesh  
[e-mail: {tahmina\_jict<sup>1</sup>, debkaushik99<sup>1\*</sup>}@cuet.ac.bd]

\*Corresponding author: Kaushik Deb

*Received April 28, 2017; revised July 2, 2017; accepted July 28, 2017;  
published November 30, 2017*

---

## Abstract

Automatic Video Surveillance System (AVSS) has become important to computer vision researchers as crime has increased in the twenty-first century. As a new branch of AVSS, baggage detection has a wide area of security applications. Some of them are, detecting baggage in baggage restricted super shop, detecting unclaimed baggage in public space etc. However, in this paper, a detection & classification framework of baggage is proposed. Initially, background subtraction is performed instead of sliding window approach to speed up the system and HSI model is used to deal with different illumination conditions. Then, a model is introduced to overcome shadow effect. Then, occlusion of objects is detected using proposed mirroring algorithm to track individual objects. Extraction of rotational signal descriptor (SP-RSD-HOG) with support plane from Region of Interest (ROI) add rotation invariance nature in HOG. Finally, dynamic human body parameter setting approach enables the system to detect & classify single or multiple pieces of carried baggage even if some portions of human are absent. In baggage detection, a strong classifier is generated by boosting similarity measure based multi layer Support Vector Machine (SVM)s into HOG based SVM. This boosting technique has been used to deal with various texture patterns of baggage. Experimental results have discovered the system satisfactorily accurate and faster comparative to other alternatives.

---

**Keywords:** Baggage, HSI model, shadow detection, occlusion detection, dynamic human body parameter, Support plane RSD-HOG, boosting similarity measure, SVM

---

A preliminary version of this paper appeared in IEEE ICECE 2016, December 20-22, Dhaka, Bangladesh. The proposed enhanced version of this paper includes shadow removal approach, mirroring algorithm for occlusion detection to enable the detection system work well in occluded environment. Further, boosting technique and addition of rotation invariance nature in HOG are also the new integration of this version. These integrations provided better detection accuracy.

## 1. Introduction

In recent years, crime in public spaces around the world has been increasing dramatically. The explosion of bombs or any other explosive substances in public places is one of the most recent crimes happening all over the world. Among some fearsome incidents, Mumbai attack in November 2008, suicidal bombing occurrence in Belgium on 22 March 2016 robbed many people's life. This crime is often done by carrying explosive substances in carried baggage and sometimes placing it in public places as unclaimed one. This crime often threatens human lives and resources. Installation of CCTV camera system in busy public and risky zones provides a tool to monitor for threat and warn the security team.

However, when the data size of CCTV footage becomes too large, it becomes difficult to monitor them manually by the security teams. These facts have motivated us to automatic video surveillance system. However, a few extensive types of research have been done in baggage detection field. In this paper, as a fresh branch of active AVSS research, a contribution is deposited to automatic baggage detection and classification field. Among some successful baggage detection system, Wahyono, Haariyono and Jo [1] extracted scalable HOG descriptor, integrated boosting strategy into SVM and developed mixture model for baggage detection. They used two human body parameters to classify baggage and set those parameters depending on the head's height. However, some of its limitations are, it fails to detect multiple pieces of baggage, misclassifies baggage when head and legs are absent or occluded and high time complexity.

However, in our previous version [22], a baggage detection framework was proposed. In this paper, we propose an enhanced version of the baggage detection framework for increasing the system performance. To improve the detection accuracy, a preprocessing module is proposed in this version. Further, we propose a support plane to make the rotational feature (RF) rotation invariance. In addition, boosting strategy is integrated into classifier to enhance the system ability when baggage is adorned by various texture patterns.

This field becomes challenging under some typical challenges like as uneven illumination condition, occlusion of objects, the absence of human parts, shadow related false alarm, different texture patterns of baggage and multiple pieces of carried baggage etc. Faced with challenges, the contributions in this paper are sixfold:

- The primary contribution of this paper is the detection of baggage having diverse texture patterns by using similarity based multilayer classifier and boosting technique. In this paper, to generate an identical feature for all baggage that is adorned by non-identical texture patterns in less time, LBP feature based multi layer SVMs have been boosted into RSD-HOG based SVM whereas mixture model in [1] introduced time complexity and high false alarm.
- Moreover, another emphasis is given to set the human body parameter dynamically and make the parameter independent of head's height which is one of the limitations of [1].
- Further, integration of rotation invariance nature into HOG using support plane dispel the narrowness of [1] and successfully detect uncarried and multiple pieces of carried baggage under the absence of human parts.
- In addition, a simple mirroring algorithm is proposed using extrema points of ROI to detect occluded individual objects and accordingly, to set body parameter correctly.

- Further, a contribution is deposited to detect shadow region and accordingly, to reduce the false alarm generated by the shadow of ROI. In this purpose, logarithmic saturation gain and projection of Y channel into I channel detect shadow region successfully.
- Furthermore, another attention is given to reduce computational time. Sliding window approach generated high computational time in [1] where background subtraction can reduce it. In this paper, a model using HSI model is proposed for background subtraction that is resistant against uneven illumination conditions.

Our preliminary version of baggage detection was first introduced in our previous work [22]. In this paper, the improvements are in four aspects: 1) preprocessing including shadow detection and human occlusion detection is added to achieve better result; 2) rotation invariance nature is added into RSD-HOG by using support plane 3) boosting strategy is integrated to generate identical feature of baggage that are adorned by different texture patterns; 4) more comparative experiments and analysis of the results are conducted.

However, to make the upcoming writing of this article more self-contained, we divide the article into several sections. Among those, Section 2 comprises of reviewing the work done in object detection including baggage detection. Section 3 addresses the proposed framework of baggage detection for video surveillance system. Section 4 presents an experimental evaluation of detection and classification accuracy of the proposed framework.

## 2. Related Work

This topic has become a subject of state-of-the-art technology in recent years. Among some successful ones, Tzanidou and Edirisinghe [2] proposed baggage detection and classification scheme by estimating moving direction of human carrying baggage and the alignment of a temporal human like template with the best-matched view specific exemplars. It integrated a unique connected baggage separation algorithm to achieve accurate baggage detection. However, they used same human body parameter scheme to classify baggage used in [1] and possesses the same limitations as in [1].

Many other effective feature descriptors used in other object detection not specifically for baggage detection are available. Among those, Venkatrayappa, Montesinos, Diep and Magnier [3] added a new dimension in feature descriptor by adding Rotational Signal Descriptor (RSD) in this field. In this purpose, they used Gaussian rotating half filter to construct feature space by finding the direction of the maximum and minimum variations of a point. The obtained descriptor showed rich, a discriminative set of local information related to the curvature of the image surface. It proved as much more efficient in feature representation of shape than any other feature descriptor. Hoang, Le and Jo [4] proposed a boosting machine using variant scale block based HOG features for pedestrian detection. They introduced a feature description using variant-scale block based Histograms of Oriented Gradients (HOG) features. In classification method, they included hybrid cascade boosting technique with adoptive boosting strategy and a Support Vector Machine.

Nigam, Deb and Khare [5] developed Moment Invariants based object recognition for different pose and appearance changes. This recognition scheme was proved as invariant against linear transformation like rotation, translation and scaling. Zhao, Liu and Lai [6] proposed multi-scale generalized index search tree (GIST) feature manifold for building recognition. They used energy spectral information that represents the spatial frequencies spread everywhere in the image. Dalal and Triggs [7] proposed Histogram of Oriented Gradient (HOG) for human detection with orientation binning of the fine scale gradients and relatively coarse spatial binning. However, HOG was sensitive to linear transformations and

got matured when it was combined with Scale Invariant Feature Transform (SIFT) approach [8]. In this approach, descriptor was calculated around the key points orientation which was resistant and invariant against noise. Thus, SIFT made HOG invariant against transformations. However, SIFT showed too much computational complexity. Abeles [9] added a new horizon in feature descriptor by using Speeded-Up Robust Features (SURF) and found SURF considerably faster than SIFT [8]. Here, orientation was estimated using Haar wavelet and then feature was computed inside a square region that aligned to the found orientation.

A modified Gradient Orientation and Location Histogram (GLOH) was proposed in [10] where initially, they combined SIFT [8] feature extraction algorithm and GLOH feature descriptor for each salient point. Finally, after salient point matching by Euclidean distance, RANSAC algorithm was used to eliminate wrong corresponding pairs. Islam, Jahan and Baek [13] considered variations in object size, orientation, illumination in object recognition. Gradient-weighted orientation histogram, the sum of wavelet responses, histogram using different color spaces combined to describe feature space of objects.

Moreover, to remove the noise like a shadow, occlusion of the interest objects, Deb, Imtiaz and Biswas [16] proposed a hybrid framework for motion region detection and an appearance based real-time motion tracking system. As a challenging branch of motion region detection, a shadow elimination process was proposed. Calculation of brightness and chromatic distortion along with the creation of projection successfully built up a shadow elimination system. In the field of shadow detection, Deb and Sunny [17] proposed a shadow detection and removal framework by using the luminance, chromatic blue, chromatic red from YCbCr color space to detect and remove shadows from images. They proposed an approach based on statistics of intensity in the YCbCr color space for detecting shadow.

An efficient background subtraction method was proposed by Tom, Haines and Tao [18]. They presented a method for background subtraction based on Dirichlet process Gaussian mixture models. This was used for estimating per-pixel background distributions. It used a nonparametric Bayesian method that allows per-pixel mode counts to be automatically inferred and avoiding over or underfitting. Vishnyakov, Sidyakin and Vizilter [19] proposed an approach for moving object detection. The approach used diffusion morphology that allows comparing images by shape matching using the projection of image one on image two. The diffusion map used complex Local Binary Pattern (CLBP) for shape matching. Finally, a background model was developed by creating projection and moving objects become visually separated from the background.

Texture analysis field has been become saturated with new interesting descriptors day by day. Amongst them, Nigam and Khare [20] used Local Binary Pattern (LBP) with its variants in recognition of human along with their activities. They integrate uniform Local Binary Pattern (LBP) with moment invariants to analyze human activity. Anamandra and Chandrasekaran [21], added a robust local descriptor named "COLOR CHILD". It removed the weaknesses of Weber Local Descriptor (WLD) [25] by considering Laplacian of Gaussian (LoG) and Tiansi fractional derivative-based filter for its orientation component.

### 3. Proposed Framework

Baggage detection and classification has become difficult job under the typical challenges. In our preliminary version [22], human occlusion made the system disable to set human parameter appropriately and so why the system failed to detect carried baggage. Furthermore, Shadow introduced many false alarms and there was no way to deal with non-identical texture patterns of baggage too. In addition, object feature was sensitive to object rotation. These

limitations decreased the detection rate. Here, we propose an enhanced version of the framework that resolves all these shortcomings and increases detection rate.

However, in this paper initially, we subtracted background instead of using a sliding window to speed up the system and took up the help of HSI model to cope up with uneven illumination condition. Then, a shadow detection approach was introduced to remove false alarm caused by shadow. Here, logarithmic saturation gain and projection of Y channel into I channel successfully detected shadow region. Then, occlusion of objects was detected using proposed mirroring algorithm to track individual objects. After that, we adopted the descriptor used in [3] since it reacts to small variation and added rotation invariance nature into it.

In the classification of baggage, the paper proposed a dynamic parameter setting approach that successfully detected multiple pieces of carried baggage under the absence of human parts. In baggage detection, a strong classifier was generated by boosting similarity measure based multi layer SVMs into HOG based SVM. This boosting strategy was used to deal with the diverse texture patterns of baggage. The proposed system structure is shown in Fig. 1.

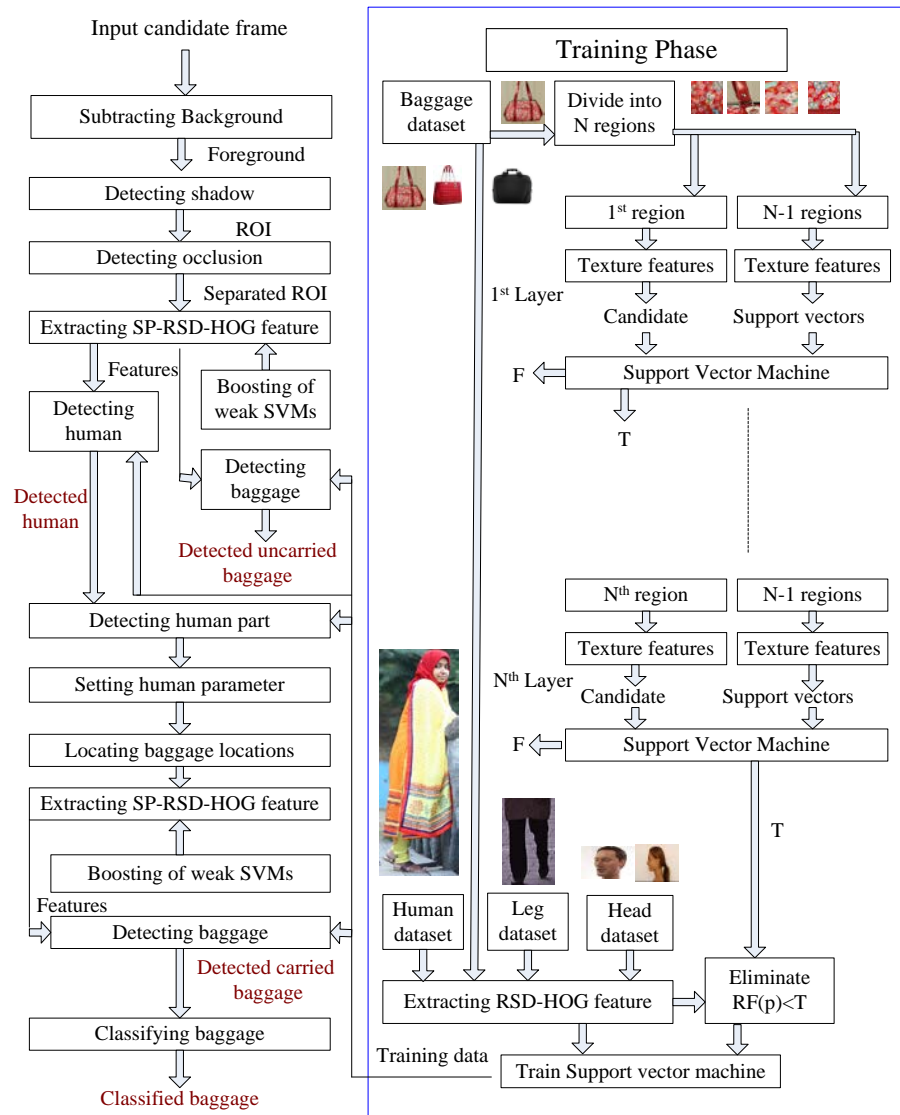


Fig. 1. Proposed baggage detection and classification framework.

The framework is divided into seven primary steps, (a) subtracting background, (b) detecting shadow, (c) detecting occlusion, (d) extracting SP-RSD-HOG feature, (e) training SVM, (f) detecting human, (g) detecting human part & setting dynamic body parameter and (h) detecting and classifying uncarried and carried baggage by boosting of SVMs.

### 3.1 Subtracting Background

In this subsection, the initial step that is the proposed background subtraction method is described. Extraction of the feature with sliding window introduces time complexity due to the large amount of candidate ROIs. Since the detection procedure was applied on video frames, background subtraction had been performed. In background subtraction, a challenging issue is different illumination condition of the input scene. To overcome this problem initially, a memory of background was built up with different illumination conditions. Then, converted them into HSI model. Among all frames, kept only those in memory which satisfies eq. (1).

$$\mu_{R(I)} > \mu_{R(H|S)} \quad (1)$$

Where,  $R(I), R(H|S) \rightarrow$  ratio between channels  $(I, H, S)$  of consecutive memory frames after applying statistical threshold. Then for an input frame, we obtained a matched memory frame. However, best match was found where memory generated the minimum value of  $\mu_{R(I)}$ . Here,  $R(I)$  is the ratio of  $I$  between input frame and memory frame. After selecting memory, for each incoming consecutive frames background was subtracted using the statistical thresholds  $T1$  &  $T2$ . The statistical thresholds are set using eq. (2). Now, if an input frame satisfies eq. (1), it denotes illumination change of input frame. For such input frame, memory frame was updated by the best matched memory of the input frame.

$$\begin{aligned} T1 &= \mu(R(I|H|S)), \text{ where } R(I|H|S) < 1 \\ T2 &= \mu(R(I|H|S)), \text{ where } R(I|H|S) > 1 \end{aligned} \quad (2)$$

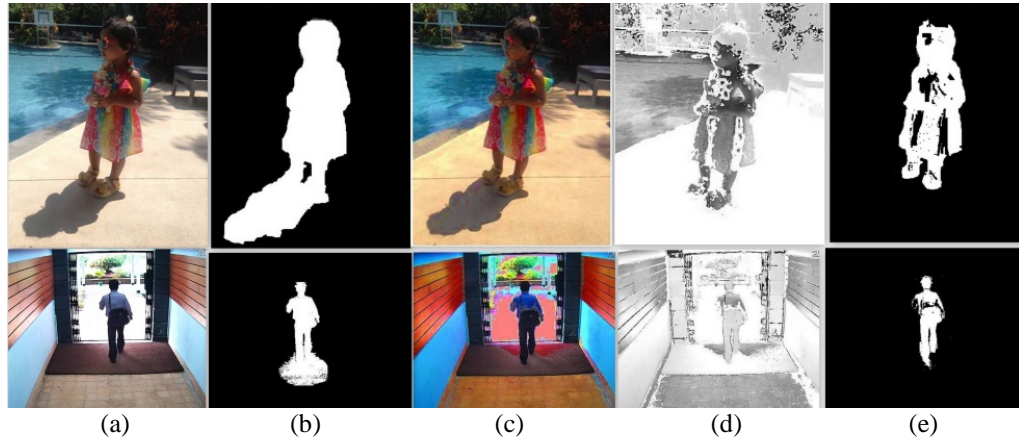
### 3.2 Detecting shadow

In this subsection, the foreground region found after background subtraction further filtered for shadow region. Object detection has become a challenging issue under different occlusion situations and the presence of similar false objects. The presence of shadow in interest region is an occlusion situation where the shadow has high similarity with the object of interest and it increases false alarm. Therefore, detection of shadow region is an important issue in object detection. In this proposed framework, an efficient way of shadow detection was unveiled using different color models. The two color models that are HSI model and YCbCr model had been used. Where,  $H, S, I$  represent hue, saturation and intensity information respectively of the image in HSI model. Besides, luminance and chrominance information of image are represented as  $Y$  and  $CbCr$  respectively in YCbCr model. However, shadow region has a unique property i.e. it has less intensity than a real black portion. The reason behind this fact is, shadow always projects on other plane and the real black portion is not. It makes the shadow lighter than a real black portion. However, intensity gain can remove the shadow from the object of interest without removing a real black portion. Here, intensity gain had been performed by creating a projection of  $Y$  channel of YCbCr model into  $I$  channel of HSI model using eq. (3). Where  $F'$  is the frame after creating projection,  $\beta$  is the controlling parameter of intensity gain ( $0 < \beta \leq 1$ ) and set as proportional to average shadow pixels intensity,  $I_{SP}$ .

$$F'_{i,j} = F_{i,j} + \beta Y_{i,j} \quad (3)$$

However, this approach generates a problem, pixels having less chromatic value fades due to intensity gain and thus lost from foreground region. However, the chromatic gain was proposed here as a solution to this problem. The chromatic gain in  $S$  channel of HSI model had





**Fig. 2.** Processing examples of shadow detection: (a) input frame, (b) background subtracted frame with shadow, (c) chromatic gain of frame, (d) projection of Y channel into I channel after chromatic gain, and (e) removal of shadow region from foreground without information loss.

been performed in the early time of intensity gain using eq. (4). Where,  $I_{SP}$  is shadow pixels intensity and  $F_{i,j}$  is the frame after chromatic gain. However, due to the logarithmic behavior, the value of  $S_{i,j}$  after chromatic gain belongs to  $<1$  when  $S_{i,j} > 0.13$ . When  $S_{i,j} \leq 0.13$ ,  $S_{i,j}$  after chromatic gain have a value  $>1$ .

$$F_{i,j} = S_{i,j} - \log(S_{i,j} + \delta), \quad \text{where, } I_{i,j} > I_{SP} \quad (4)$$

In this case,  $\delta$  is used for normalizing the greater than 1 value into less than 1. The lowest value of  $\delta$  that can normalize the value of  $S_{i,j}$  is 0.1. Therefore, set  $\delta = 0.1$ . **Fig. 2 (c)** visualizes the frame after chromatic gain whereas **Fig. 2 (d)** visualizes the frame after creating projection on highly chromatic frame. Then shadow pixels were removed from the background subtracted foreground using eq. (5), where,  $F_{i,j}$  is the foreground pixel,  $I'_{i,j}$  is the I channel's value after intensity gain,  $S'_{i,j}$  is the S channel's value after chromatic gain,  $NSp$  is non-shadow pixel and  $Sp$  is shadow pixel. The equation stated that, which pixels are faded after chromatic and intensity gain are shadow pixels and others that not faded are non-shadow pixels. The outcome of this procedure is verified in **Fig. 2 (e)**.

$$\begin{cases} F_{i,j} = NSp, & I'_{i,j} \leq 255 \text{ \& } S'_{i,j} \gg 0 \\ F_{i,j} = Sp, & I'_{i,j} = 255 \text{ \& } S'_{i,j} \cong 0 \end{cases} \quad (5)$$

### 3.3 Detecting Occlusion

After getting the ROIs without shadow, a mirroring algorithm was applied to unroll the overlapping or occlusion of interest objects. Here, we used the global and local extrema to make the mirror. The basic advantage of this algorithm is, it can detect occlusion incident and occluded region without prior knowledge of separated individual objects in the scene view. However, the algorithm simulation steps are viewed in **Fig. 3** whereas **Fig. 4** explores some simulation results. In **Fig. 3 (b & c)**, operational statements from step 2-5 of the mirroring algorithm are executed. Here in step 5, Next\_area was set to left in the first iteration. Then we executed step 6 i.e. mirrored left portion to right w.r.t.  $L_1$ . Then, both portions were labeled with green color as shown in **Fig. 3 (d)**. Then according to step 7 and 1<sup>st</sup> if, we labeled the 2<sup>nd</sup> region with red color as shown in **Fig. 3 (e)**. Finally, as shown in **Fig. 3 (f)** overlapped region was detected where two colors overlapped as stated in step 7, 2<sup>nd</sup> if.

---

Mirroring algorithm for occlusion detection

---

**Step1:** Find extrema points  $E_{\max}(x_n, y_n)$  and  $E_{\min}(x_m, y_m)$ ,  $n \rightarrow$  number of maximum extrema,  $m \rightarrow$  number of minimum extrema, among all boundary points of a ROI  $\{B(x, y), \text{ where } (x, y) \rightarrow 1, 2, \dots, N\}$ ,  $N$  is the number of boundary points

**Step2:** Find global maxima point  $G_{\max}(x, y) \in E_{\max}(x_n, y_n)$  (if more than one, select the floor(middle) one)

**Step 3:** Set a line  $L_1$  from  $G_{\max}(x, y)$  to downward crossing the ROI

**Step 4:** Calculate area of the ROI on both side of  $L_1$ , Resulting,  $\text{Area}(L_{1, \text{left}}, L_{1, \text{right}})$

**Step 5:** if  $\text{Area}(L_{1, \text{left}}) < \text{Area}(L_{1, \text{right}})$  ( $1^{\text{st}}$  iteration only)

Next\_area  $\rightarrow$  left

else

Next\_area  $\rightarrow$  right

or

if  $(G_{\max}(x, y) > \text{recent } G_{\max}(x, y))$  (for later iteration)

Next\_area  $\rightarrow$  left

else

Next\_area  $\rightarrow$  right

**Step 6:** if (Next\_area  $\rightarrow$  left)

1. Set a color  $C_1$  to Region( $L_{1, \text{left}}$ , upto next maxima) & Mirror Region( $L_{1, \text{left}}$ , upto next maxima)  $\rightarrow$  Region( $L_{1, \text{right}}$ )

2. Set the same color  $C_1$  to Mirrored Region ( $L_{1, \text{right}}$ )

3. Set traverse  $\rightarrow$  clockwise ( $1^{\text{st}}$  time iteration only)

else

1. Set a color  $C_1$  to Region( $L_{1, \text{right}}$ , upto next maxima) & Mirror Region( $L_{1, \text{right}}$ , upto next maxima)  $\rightarrow$  Region( $L_{1, \text{left}}$ )

2. Set the same color  $C_1$  to Mirrored Region( $L_{1, \text{left}}$ )

3. Set traverse  $\rightarrow$  Anti-clockwise ( $1^{\text{st}}$  time iteration only)

**Step 7:** if (non\_colored ROI > 20% of ROI & next  $L_{\max}(x, y) \neq G_{\max}(x, y)$ )

1. Find Local extrema with traverse direction & get next  $L_{\max}(x, y) \in E_{\max}(x_n, y_n)$  among  $B(x, y) \notin G_{\max}(x, y)$ .

3. Set  $G_{\max}(x, y) = L_{\max}(x, y)$  &  $C_1 = C_2$ , goto step 3

else

if (non\_colored ROI < 20% of ROI)

1. Label with a new color  $C_m$  where two colors overlapped

or

1. If one color exist no occlusion occurred

2. Exit;

else

if (non\_colored ROI > 20% of ROI & next  $L_{\max}(x, y) == G_{\max}(x, y)$ )

1. Find Local extrema with traverse direction & get next  $L_{\min}(x, y) \in E_{\min}(x_n, y_n)$  among  $B(x, y)$

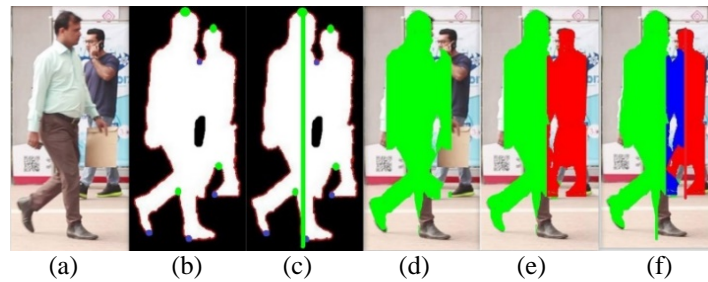
3. Set  $G_{\min}(x, y) = L_{\min}(x, y)$  &  $C_1 = C_2$ , goto step 3

---

### 3.3 Extracting SP-RSD-HOG Feature

After getting the separated ROIs, the feature was extracted from ROIs. As shown in [Fig. 5 \(a\)](#) at pixel  $(x, y)$ , a derivative kernel was applied to obtain a derivative information  $\varphi(x, y, \theta)$  in a function of orientation  $\theta \in [0; 360^\circ]$ . Note that, it scalps the hazy effect generated in isotropic filters by making the features sustainable along an individual direction.



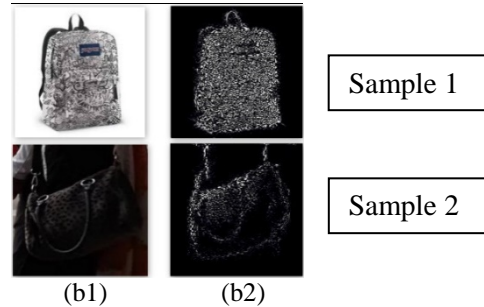
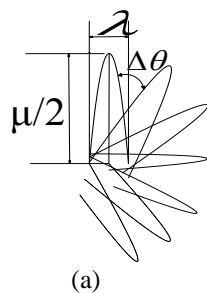


**Fig. 3.** Steps of occlusion detection: (a) input frame, (b) extrema of extracted ROI (green circle→maxima points, blue circle→minima points), (c) line  $L_1$  from the global maxima point, (d) mirroring left portion to right w.r.t.  $L_1$  and label with green color, (e) mirroring w.r.t. next extrema and label with red color as stated in step 7 and 1<sup>st</sup> if, and (f) detected overlapped region using step 7 and 2<sup>nd</sup> if (label with blue color).



**Fig. 4.** Sample simulation results of occlusion detection (blue portions are occluded region).

Further, the filter preserves the slightest variation along a direction and thus can detect baggage from the approximate same colored background. The derivative kernel is defined in eq. (6). Where  $I \rightarrow$  image,  $\theta \rightarrow$  orientation of the filter  $[0; 360^\circ]$ ,  $C \rightarrow$  normalization coefficient,  $(x, y) \rightarrow$  pixel coordinates and  $(\mu, \lambda)$  the standard deviation of the anisotropic Gaussian kernel called height and width of the filter. Only the casual part of this filter along the Y axis used. This is obtained by cutting the kernel in the middle. This is an operation that corresponds to the Heaviside function  $H$ .



**Fig. 5.** Description of filter & application on frame: (a) a thin rotating Gaussian derivative half-filter applied to a key point  $(x, y)$ , (b1) input frame, and (b2)  $||\nabla I||$  frame after applying the defined filter.

Thus Rotational-Signal-Descriptor (RSD) had been obtained by rotating the filter around a point in steps of  $5^\circ$  here. The step was selected  $5^\circ$  because increasing the rotation step results in loss of information. We computed the gradient  $||\nabla I||$  and the two angles for the pixel at location  $(x, y)$  by considering the global extrema of the function  $\varphi(x, y, \theta)$ . Two of these global extrema were combined to maximize  $||\nabla I||$  shown in eq. (7). The resulting RSD-HOG was generated using two angles maxima  $\theta_1$  and minima  $\theta_2$ . Finally, bin the two angles separately as  $F_1$  and  $F_2$  for the defined blocks. Merging of these two angles as in eq. (8) resulted a robust description. Where,  $N$  is the total number of block and  $N_1$  is the number of block along the row of interest object. **Fig. 5 (b2)** visualizes the frame after applying rotating filter where sample 1 represents a baggage with diverse texture pattern and sample 2 represents a baggage belongs to an approximatly same colored background.

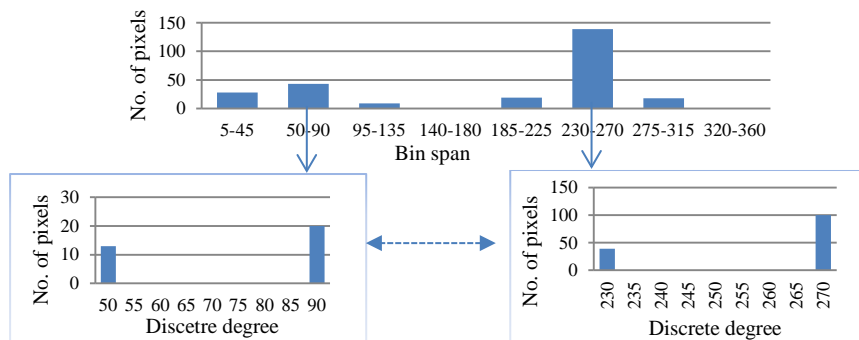
$$\varphi(x, y, \theta) = I * C.H(-y).x.e^{-\left(\frac{x^2}{2\lambda^2} + \frac{y^2}{2\mu^2}\right)} \quad (6)$$

$$\begin{cases} \theta_1 = \arg \max_{\theta \in [0, 360^\circ]} Q(xp, yp, \theta) \\ \theta_2 = \arg \min_{\theta \in [0, 360^\circ]} Q(xp, yp, \theta) \\ ||\nabla I|| = \max_{\theta \in [0, 360^\circ]} Q(xp, yp, \theta) - \min_{\theta \in [0, 360^\circ]} Q(xp, yp, \theta) \end{cases} \quad (7)$$

$$RSD\_HOG = \begin{bmatrix} F_1 \begin{bmatrix} \theta_1 bins_{block 1} & \cdots & \theta_1 bins_{block N_1} \\ \vdots & \ddots & \vdots \\ \theta_1 bins_{block N-N_1} & \cdots & \theta_1 bins_{block N} \end{bmatrix} \\ F_2 \begin{bmatrix} \theta_2 bins_{block 1} & \cdots & \theta_2 bins_{block N_1} \\ \vdots & \ddots & \vdots \\ \theta_2 bins_{block N-N_1} & \cdots & \theta_2 bins_{block N} \end{bmatrix} \end{bmatrix} \quad (8)$$

SVM provides a good detection result when the object aligned with zero rotation. However, the performance of SVM is decreased with object rotation. An accurate detection system requires rotation invariance nature of the descriptor. To achieve this goal, we provided a plane for each bin and isolated each discrete value by counting those separately as shown in **Fig. 6**.

However, in case of RSD-HOG, the effect of object rotation on  $\theta_1$  is described in **Fig. 7**. When the object aligned normal, the angle was found  $\theta_1$  by using the rotational filter. When the object rotates, the angle was modified to  $\theta_1'$  as shown in **Fig. 7 (b)**. The change  $\Delta\theta_1 = \theta_1 \sim \theta_1'$  replaced  $\theta_1$  into another bin. Now, the crucial point is if we could anti rotate  $\theta_1'$  by an amount of  $\Delta\theta_1$ , then  $\theta_1'$  will be re-placed into  $\theta_1$  again. The anti-rotation operation was performed by swapping the bin values. In swapping of bin values, a critical issue arise. Let, one bin had discrete values of  $\{5, 10, 15, 20, 25, 30, 35, 40, 45\}$ . Now the anti rotation of  $5^\circ$  swapped those pixels into previous bin which had value of  $5^\circ$ . The support plane provided the information about pixel amount of discrete degrees in a bin as shown in **Fig. 6**.



**Fig. 6.** Proposed support plane for RSD-HOG (SP-RSD-HOG).



**Fig. 7.** Effect of object rotation: (a) object without rotation, and (b) object with rotation,  $\theta_1$  changes to  $\theta'_1$  with a rotation angle  $\Delta\theta_1$ .

In addition, when object is rotated by  $20^\circ$  i.e.  $20^\circ \leq \Delta\theta_1 \leq 340^\circ$ , block sequences in feature space are also rotated notably. To regain the block sequences initially, we rotated blocks of feature space i.e. feature matrix defined in eq. (8) by  $\Delta\theta_1$ . Here,  $F_1$  and  $F_2$  were rotated separately. After that, bin values were swapped.

### 3.4 Training SVM

After extracting features, we trained up the classifier with appropriate training data. Training data are the most considerable part of a detection system since relevant training data provided an efficient detection system. In this work, RSD-HOG based SVM used radial basis function kernel and trained up with interest objects in its 4 z-axis orientation since each object of interest varies with its orientation and appearances. In the case of baggage detection, different appearances of baggage misguided the weak RSD-HOG features based single layer SVM classifier. To make an efficient baggage detection system, texture feature based multi layer SVMs were boosted into RSD-HOG features based single layer SVM.

Generally, baggage possesses common texture patterns that cover whole baggage. Here, Local Binary Pattern (LBP) had been used to analyze the baggage texture. In this purpose initially, one interest object was divided into  $N$  regions  $\{R_1, R_2, \dots, R_N\}$ . Then, LBP features  $LF_i$  were extracted from each  $R_i$  and a SVM layer was built up for each region consecutively. Thus we constructed a multilayer ( $N^{\text{th}}$  layer) SVM. In each layer, LBP texture feature,  $LF_C$  of a region  $R_i$  as a candidate region was analyzed subject to the other region's  $\{R_{1-n}, \notin R_i\}$  LBP texture features,  $LF_{TD}^{n-1}$  as training data of SVM. **Fig. 8** visualizes the similarity texture based multi layer SVM.

Here, distance metric as defined in eq. (9) was used as kernel function for SVMs where  $L$  is the length of the LBP feature vector. Each layer resulted a positive output if texture is similar to the other region's texture otherwise gave a negative result. The SVM layers which gave positive result had been boosted into RSD-HOG based weak classifier to eliminate rotational feature, RF (gradient) of that region. After elimination of sort gradients generated due to the pattern of baggage, the resulted RSD-HOG described the baggage characteristics more uniquely and clearly as shown in **Fig. 9**.

At this stage, boosting of multilayer SVM generated a strong RSD-HOG feature-based classifier for baggage detection. The portioning scheme of a training data was chosen depending on the center point of the object. Traversing to the eight directions from the center until reaching the boundary gave  $N$  regions. The region size was set with respect to the height and width of the object.

$$DM(TD^{n-1}, C) = \sum_{j=1}^L \left| \frac{LF_{TD}^{n-1}(j) - LF_C(j)}{1 + LF_{TD}^{n-1}(j) + LF_C(j)} \right| \quad (9)$$

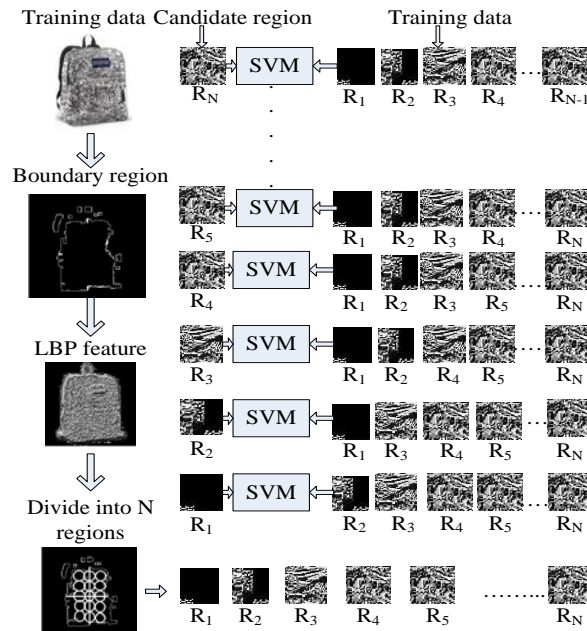


Fig. 8. Texture based multi layer SVM.

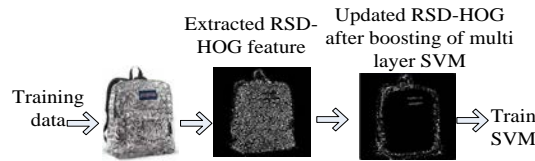


Fig. 9. Boosting of multi-layer SVMs.

### 3.5 Detecting Human

After training the classifier, in this phase, SP-RSD-HOG features of candidate ROIs were fed into RSD-HOG based SVM without boosting for human detection. Since carried baggage detection & classification depends on the human parameter setting so, human detection is the most important part to direct the system in a correct way. Here, the resized size  $M \times N$  was set to  $256 \times 256$  since extra enlargement lost salient features and too much small size introduced misconception about salient features.

---

#### Human detection process

---

for  $i=1$  to each ROIs

**Step 1:** Resize input Region of Interest into  $M \times N$  size

**Step 2:** Divide into  $16 \times 16$  blocks

**Step 3:** Extract SP-RSD-HOG features from each block & build up feature vector

**Step 4:** Execute detection phase using single layer RSD-HOG based SVM

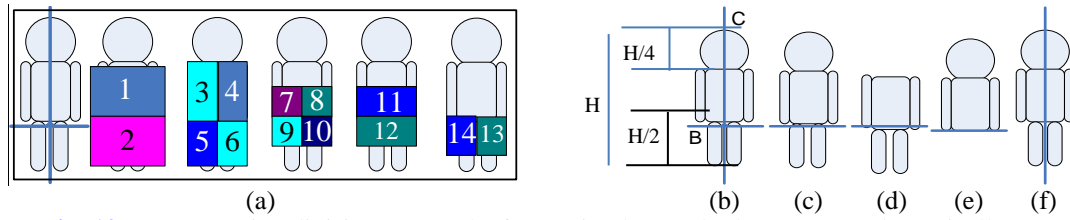
Decision: Human/non-human

end for

---

### 3.6 Detecting Human Part and Setting Dynamic Body Parameter

After human detection, in this stage, we set some human parameter dynamically. Here, two primary body parameters were employed that are, bend line  $B$  and vertical line  $C$  that traversing towards the center of the human horizontally and vertically respectively as shown in Fig. 10 (b). However, an approach was proposed here to set the bend line  $B$  dynamically. The



**Fig. 10.** Human region division approach after setting human body parameter dynamically: (a) fourteen regions of detected human, (b) human body parameter, (c) set B when all parts are present, (d) set B in absence of head, (e) set B in absence of leg, and (f) set vertical line, C.

approach made the system able to fix the position of the bend line  $B$  although head or leg was occluded as shown in **Fig. 10 (d & e)**. In this purpose initially, two portions of the detected human, upper  $H/4$  portion and lower  $H/2$  portion, were tested with RSD-HOG based SVM to detect head and leg in those portions respectively. Where  $H \rightarrow$  height of bounding box. Then, the positions of the vertical & bend line were fixed using eq. (10) Where,  $T$  is the top position of the bounding box,  $W$  is the width of the bounding box, +ve denotes object is present and -ve means absent .

$$\begin{cases} B = T + \frac{7H}{10}; & \text{if head, leg} \rightarrow +ve \\ B = T + H; & \text{if head} \rightarrow +ve, \text{leg} \rightarrow -ve \\ B = T + \frac{H}{2}; & \text{if head} \rightarrow -ve, \text{leg} \rightarrow +ve \end{cases} \quad (10)$$

$$C = \frac{W}{2}$$

After setting human body parameter, to fix the possible locations of the carried baggage, we divided the human region based on bend line and vertical line into fourteen regions. The possible locations of carried baggage can belong to these 14 regions only. **Fig. 10 (a)** demonstrates the region division scheme.

### 3.7 Detecting and Classifying Uncarried and Carried Baggage by Boosting of SVMs

Finally, the remaining job is to detect and classify carried and uncarried baggage. Uncarried baggage is placed as unclaimed at anywhere of a public place. In the case of carried baggage generally, which is held in hand ordinarily placed in below of the bend line and which are carried in shoulder placed upon the bend line. However, according to this scenario, in this work carried baggage were classified into two categories like in **Fig. 11**.

However, in this phase, the extracted SP-RSD-HOG feature from candidate ROIs and fourteen carried baggage locations were refurbished. Initially, the extracted LBP feature from candidate ROIs and fourteen carried baggage locations were fed into multi layer SVMs. Then we boosted the result of multilayer SVMs into RSD-HOG feature based SVM as described in the training phase.



**Fig. 11.** Classification of carried baggage into two categories: (a) category 1: backpack residue up on bend line (green color box), and (b) category 2: handbag residue below the bend line (blue color box).



---

Baggage detection & classification process

---

for  $i=1$  to each ROIs

**Step 1:** Resize input Region of Interest into  $M \times N$  size & divide into  $16 \times 16$  blocks

**Step 2:** Extract SP-RSD-HOG and LBP features of ROI

**Step 3:** ROI is divided into  $N$  regions as shown in Fig. 8

- Apply multi-layer LBP based SVMs to each region like Fig. 8
- Boosting results into RSD-HOG based SVM by updating features like in Fig. 9
- Execute detection phase
- Output: Detected baggage, DB

**Step 4:** If DB is found in fourteen carried locations

Classify baggage into categories based on bend line

end for

---

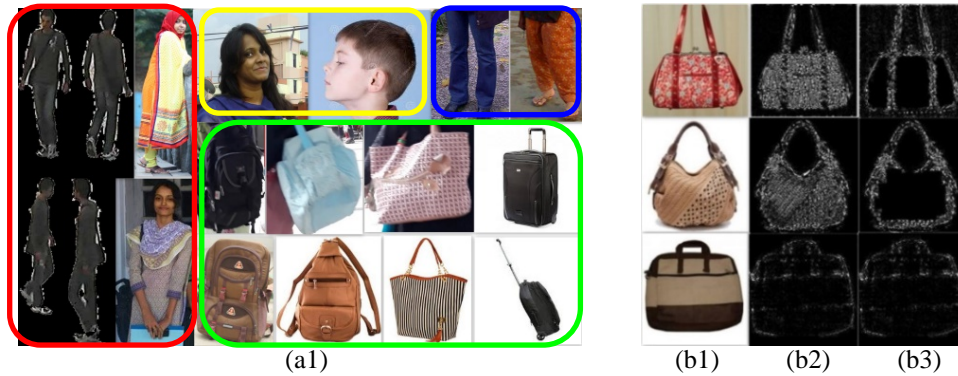
#### 4. Experimental Results and Analysis

The framework was tested on our own data set and on some standard data sets together with 540 iLids frames and 141 CAVIAR video frames. Our dataset was collected from “komolapur railway station,” “bashundhara city” at Dhaka, Bangladesh and some CCTV footage of CUET campus were collected from the authority. In addition, some videos were captured randomly in different places. However, some sample dataset with all the considered challenges of the proposed framework is given in Fig. 12. Initially, every interest object was rescaled to  $256 \times 256$ . Then, divided the ROI into  $16 \times 16$  blocks and found 256 blocks. After that, each block was bin with a bin span of  $45^\circ$  and thus resulting 8 bin for each block. The size of the block was set to  $16 \times 16$  since the filter rotation with  $5^\circ$  about a pixel can understand the changes in this portion only. However, the generated feature space size for  $\theta_1$  was found,  $8 \times 256 = 2048$  bin. Finally, for both  $\theta_1$  and  $\theta_2$ , feature space size become  $2048 \times 2 = 4096$  bin. The framework had been simulated in MATLAB simulator with corei5 processor and 4GB ram. The classifier was trained up with four types of interest object as shown in Fig. 13 (a1). The data sizes were about 510 human, 870 baggage, 200 head and 160 leg images. Moreover, training baggage data are shown in Fig. 13 (b1) where Fig. 13 (b3) visualizes associate training data after boosting.



**Fig. 12.** Testing dataset: (a) baggage without human, uneven illumination, (b) human part absent, (c) approximately same background color, (d) varied distance, (b,e) human occluded, with shadow, multiple pieces of carried baggage, non-identical texture of baggage, and (f) baggage rotated.





**Fig. 13.** Training data: (a1) four types of objects, (b1) baggage dataset, (b2) extracted RSD-HOG features of baggage dataset, and (b3) updated RSD-HOG feature of baggage after boosting of SVMs.

**Table 1** summarizes the result of testing data of human, carried and uncarried baggage with different texture pattern under normal situation, different illumination, occlusion of human portions, shadow effect, object rotation and the situation when multiple pieces of baggage were carried. Visually, **Fig. 14** describes the simulation data in each step of the proposed framework. The detection & classification success under the presence of shadow, human occlusion, non-identical texture pattern of baggage and multiple pieces of carried baggage is shown in **Fig. 14**. **Fig. 15 (a)** visualizes the detection and classification success of handbag under occlusion of human parts which was the limitation of [1]. Furthermore, the proposed system detected baggage surrounded by approximately same color successfully as shown in **Fig. 15 (b)**.

**Table 1.** Experimental results

Dataset	Situation	Testing	Detected	Classified
Human	Normal	898	852	-
	Illumination changed	361	335	-
	Shadow effect	1022	1002	-
	Occluded	113	110	-
Backpack	Normal	861	750	706
	Illumination changed	223	195	183
	Occluded	116	101	95
	Rotated	10	8	8
Handbag	Normal	992	864	814
	Illumination changed	237	207	195
	Occluded	51	45	42
	Rotated	11	10	10
Backpack & handbag	Multiple baggage	141	123	116
Uncarried baggage	Normal	95	90	-
	Illumination changed	50	38	-
	Shadow effect	53	50	-
	Rotated	7	7	-



**Fig. 14.** Experimental steps of proposed framework under challenging situation (with shadow, occlusion of human, non-identical texture pattern and multiple pieces of carried baggage): (a1) input frame, (a2) background subtracted frame with shadow, (a3) after removing shadow region, (b1) input occluded ROI, (b2, b3 & b4) detecting occlusion in ROI using mirroring algorithm described in subsection 3.3, (b5) separated ROIs, (c1) input separated ROIs, (c2) extracted RSD-HOG of ROIs, (c3) detected humans using single layer RSD-HOG based SVM, (d1) set human body parameter and divide human into 14 regions, (d2) extracted RSD-HOG from regions (shown 5<sup>th</sup> and 13<sup>th</sup> region only), (d3) RSD-HOG after boosting multi layer SVMs, and (d4) detected baggage and classified as handbag.



**Fig. 15.** Analysis of detection & classification success under difficult scenarios: (a) occlusion of human part (head absent), and (b) baggage with the approximately same colored background.

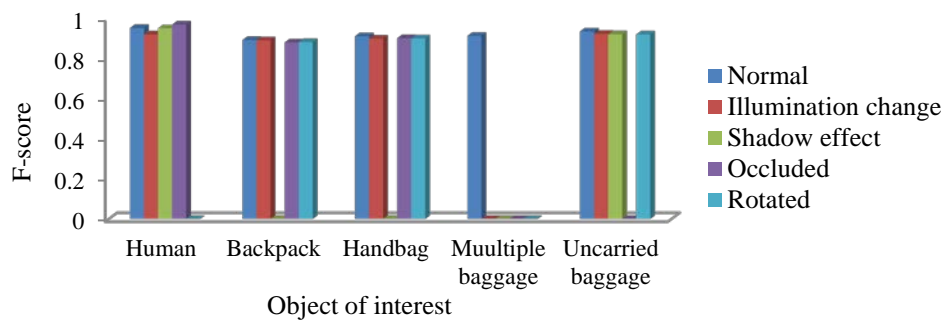
**Fig. 17 (a)** shows the detected sequence of the frame of the iLids data set where a human left his bag into public place. Whereas, **Fig. 17 (b)** shows frames where human occlusion took place. In addition, **Fig. 17 (c)** visualizes the framework success when the object was rotated and at the same time **Fig. 17 (d)** shows uncarried baggage detection without any human in uneven illumination condition. Furthermore, some random experimental results are given in **Fig. 18**. The average detection accuracy of the proposed system for 5241 frame samples of 20 videos in a different environment was found about 91.34% and classification accuracy was found about 85.32%. These results are compared with [1], [2] and [13] in terms of detection and classification accuracy and time complexity in **Table 2** and found the proposed system has superior performance than [1], [2] and [13].

We further evaluated F-score used in [23], a measure of test accuracy. It considers both the precision  $p$  and the recall  $r$  of the test to compute the score. This is the harmonic mean of the precision  $p$  and recalls  $r$  as defined in eq. (11). we figure out the F-score for testing the detection accuracy of human, backpack, handbag, multiple pieces of baggage and uncarried baggage in different occlusion situation in **Fig. 16**.

$$F = 2 \cdot \frac{p \cdot r}{p + r} \quad (11)$$

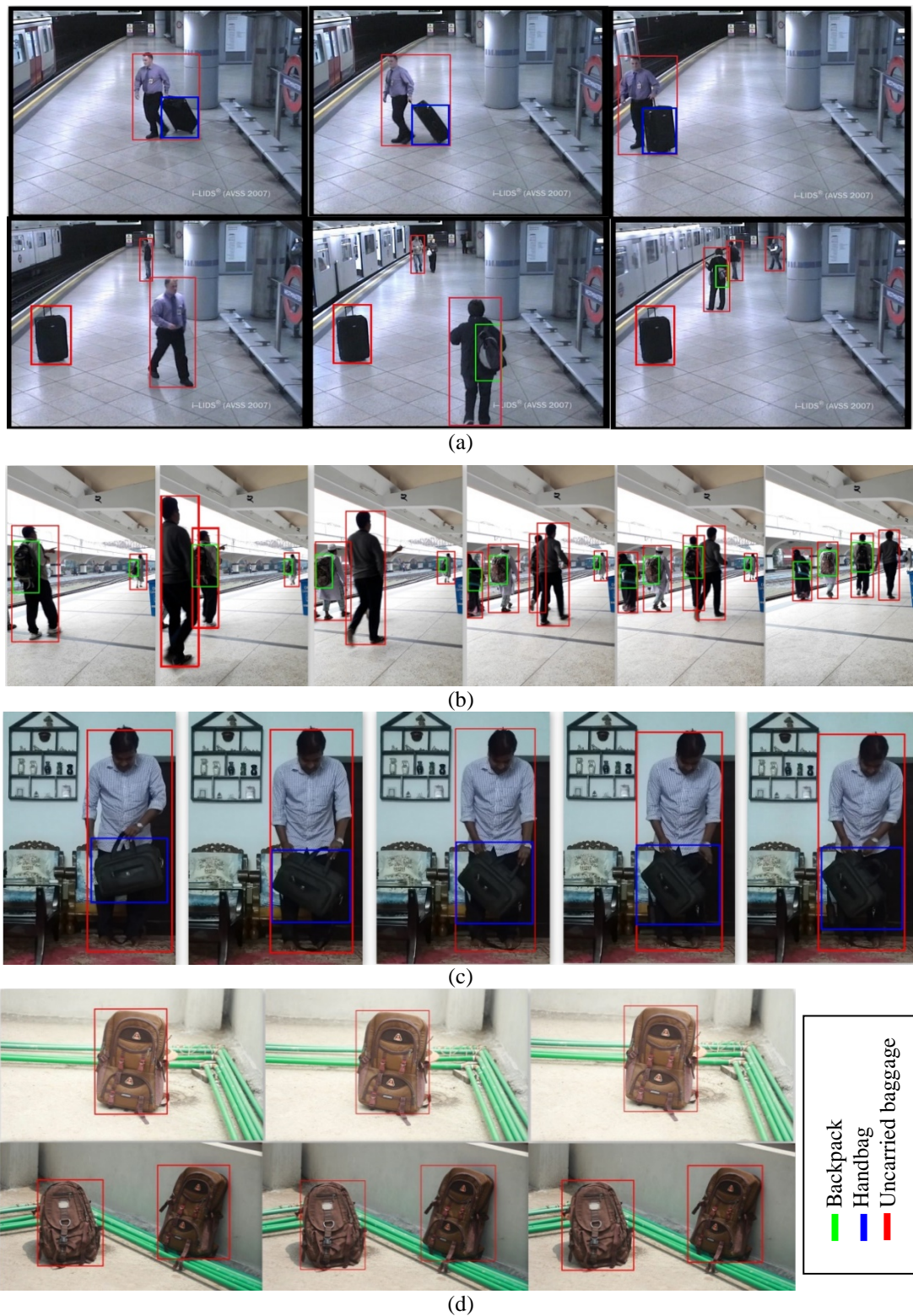
**Table 2.** Performance evaluation of the proposed framework in terms of accuracy and time complexity

	Detection accuracy	Classification accuracy	Average computational time (s)
[1]	77.02%	59.40%	3.86
[2]	60.78%	-	-
[13]	70.16%	60.05%	2.10
The proposed method	91.34%	85.32%	1.00



**Fig. 16.** F-score for different types of interest object under various conditions.



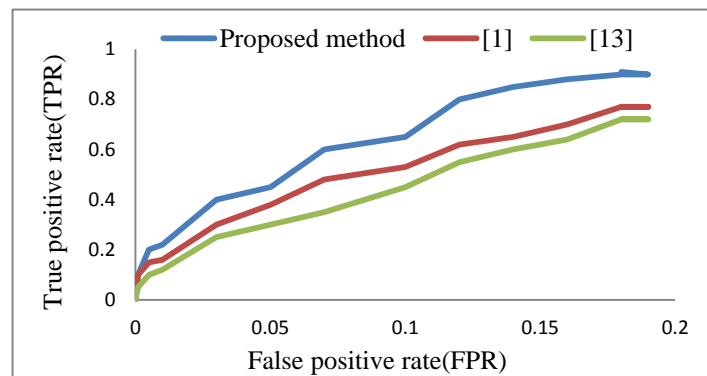


**Fig. 17.** Sequence of detected frames: (a) from iLids dataset (human is leaving baggage), (b) frames with human occlusion, (c) frames with object rotation, and (d) baggage detection without human in uneven illumination condition.



**Fig. 18.** Experimental results of detected and classified human and baggage on the own data set.

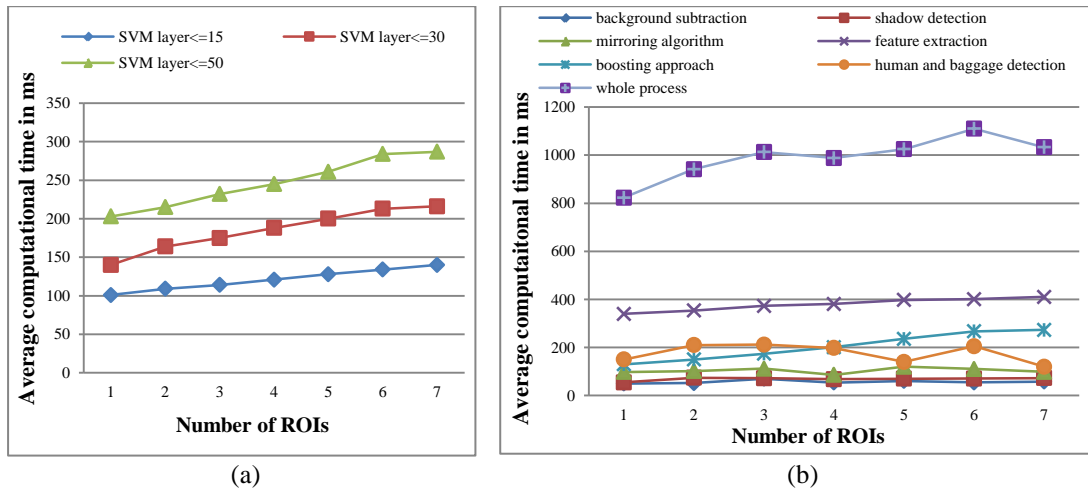
In addition, the performance of the proposed system has measured and compared with [1] and [13] by ROC curve in Fig. 19. Furthermore, some performance evaluation matrix defined in Table 3 are evaluated to test accuracy where accuracy is calculated as defined in [24].



**Fig. 19.** ROC curve of proposed method, [1] and [13] for our own data set.

**Table 3.** Comparative performance analysis of the proposed framework in different evaluation matrix

On total 5241 frames	Sensitivity/ TPR/ Recall	Precision	Specificity	FPR	Accuracy(%)
	TP/(TP+FN)	TP/(TP+FP)	TN/(TN+FP)	FP/(FP+TN)	(TP+TN)/ (TP+FP+FN+TN)
<b>Proposed method</b>	0.9151	0.9071	0.8252	0.1732	88.27%
<b>[1]</b>	0.7789	0.7310	0.6956	0.3043	73.85%
<b>[2]-claimed</b>	0.7750	0.7381	-	-	60.78%
<b>[13]</b>	0.7215	0.6902	0.6203	0.3105	69.31%



**Fig. 20.** Analysis of time complexity: (a) required time for multi layer SVMs, and (b) computational time of each step and whole system framework.

The required computational time of the proposed system to process a frame is analyzed in Fig. 20. Since the framework has several steps, Fig. 20 (b) represents the average required time for all steps. The time complexity of all steps seems to be constant for a frame and it refers to  $O(1)$  time complexity except boosting approach. In Fig. 20 (a), a detailed map of time complexity is shown for boosting of multilayer SVMs. It represents that time complexity gradually increased with the increase of ROIs and SVM layers. If  $m$  refers to the number of SVM layers and  $n \rightarrow$  number of ROIs then multilayer SVMs required  $\alpha m \times \beta n$  time where,  $\alpha, \beta$  represent non linearity.

However, SVM layers do not cross the limit of 50 since they were built up related to the height and width of the ROI. So,  $m$  tends to be a constant and remaining time complexity refers to  $O(\beta n)$  or  $O(n)$  complexity. However, since the number of ROIs were not increased to too much high value in a scene view and the increase of elapsed time w.r.t  $n$  was not so higher,  $O(n)$  complexity can be ignored. Finally, the average computational time of the whole process is mapped in Fig. 20 (b) and it was found tends to be constant time complexity.

The above experimental results and evaluation proved our proposed system outperform the existing successful one in terms of accuracy and computational time. The superior detection accuracy was found due to rotation invariance rotational descriptor. Furthermore, removal of shadow increased precision and decreased FPR by reducing the false alarm. In addition, boosting strategy and occlusion incident detection increased the TP i.e. recall rate by generating identical feature of non-identical texture patterns. Moreover, better classification accuracy was found because of dynamic human body parameter setting approach. Also, the system has been proved faster in terms of time complexity since it performed background subtraction with illumination changing challenge instead of using a sliding window.

## 5. Conclusion

In recent years bombing attack by terrorist spreads all over the world and they basically use baggage to carry explosive substances. In this paper, a framework is proposed for baggage detection with the vision of ensuring security in public places by monitoring human and baggage along with or placed as unclaimed. However, the outcome of the proposed framework is detected baggage. Further, a casual scanner like laser scanner can easily scan the materials



inside it.

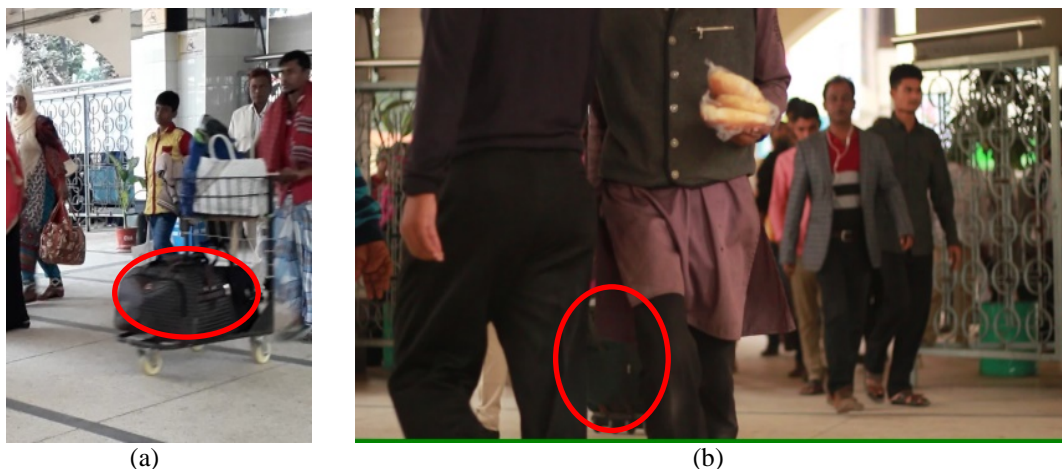
However, in the proposed framework initially, background subtraction is performed instead of using sliding window approach to speed up the system. Sliding window approach generates a large amount of ROIs during sliding the window over the frame. This is why it generates huge ROIs and increase time complexity by processing each ROI one after another. However, HSI model is used in background subtraction to cope up with different illumination condition.

Then, an efficient shadow detection framework is introduced using HSI and YCbCr model to remove false alarms that are generated by shadow. It employs logarithmic saturation gain and projection of Y channel into I channel to successfully detect shadow region. This shadow detection approach can detect any type of shadow including soft, hard and self-shadow. Then, occlusion is detected to separate individual ROIs. In this purpose, we propose a mirroring algorithm to track individual objects.

After that, rotational signal descriptor with support plane (SP-RSD-HOG) is extracted to make the descriptor rotation invariance. After training the SVM classifier, we detect human with it. Then, dynamic approach is evaluated to set human body parameter appropriately. This approach set the parameters depending on the presence of the human parts instead of depending on head's height as in [1]. Finally, in baggage detection, we boost the similarity measure based multi layer SVMs and generate a strong classifier to deal with different texture patterns of baggage.

After the experiment, the proposed system shows a satisfactory detection & classification rate with multiple pieces of baggage detection and classification. In other words, experimental results have given evidence that [1], [2], [13] perform inferior to the proposed method since the proposed framework overcomes some of their limitations. However, one of the limitations of the framework is, it fails to detect bundle of baggage that resides together as shown in Fig. 21 (a). In that case, proposed mirroring algorithm has failed to track individual bag since those do not have extrema points. Moreover, a bundle of baggage does not represent feature appropriately.

In addition, another limitation is, it fails to detect baggage in a too much-occluded situation where objects are not visually separable. In such occluded situation like in Fig. 21 (b) the framework fails to set the appropriate possible locations of baggage. Besides, features remain hidden due to the occlusion too. However, possible promotion of these limitations can improve the system performance in a notable form.



**Fig. 21.** Failure aspects of the proposed framework of baggage detection: (a) bundle of baggage, and (b) failed to set baggage locations.

## References

- [1] Wahyono, J. Haariyono, and K.H. Jo, "Body part boosting model for carried baggage detection and classification," *Neurocomputing*, vol. 228, pp. 106-118, March 2017. [Article \(CrossRef Link\)](#).
- [2] G. Tzanidou and E. A. Edirisinghe, "Automatic baggage detection and classification," in *Proc. of IEEE International Conference on Intelligent Systems Design and Applications*, pp. 825-830, 2011. [Article \(CrossRef Link\)](#).
- [3] D. Venkatrayappa, P. Montesinos, D. Diep and B. Magnier, "RSD-HoG: a new image descriptor," *ser. Lecture Notes on Computer Science*, Springer International Publishing Switzerland, vol. 9127, 2015. [Article \(CrossRef Link\)](#).
- [4] V. D. Hoang, M. H. Le and K.H. Jo, "Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection," *Neurocomputing*, vol. 135, pp. 357-366, 2014. [Article \(CrossRef Link\)](#).
- [5] S. Nigam, K. Deb and A. Khare, "Moment invariants based object recognition for different pose and appearances in real scenes," in *Proc. IEEE International Conference on Informatics, Electronics & Vision*, pp. 1-5, 2013. [Article \(CrossRef Link\)](#).
- [6] C. Zhao, C. Liu and Z. Lai, "Multiscale GIST feature manifold for building recognition," *Neurocomputing*, vol. 74, pp. 2929-2940, 2011. [Article \(CrossRef Link\)](#).
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society*, vol. 1, pp. 886-893, 2005. [Article \(CrossRef Link\)](#).
- [8] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004. [Article \(CrossRef Link\)](#).
- [9] P. Abeles, "Speeding up SURF," *ser. Lecture Notes on Computer Science*, Springer-Verlag Berlin Heidelberg, vol. 8034, 2013. [Article \(CrossRef Link\)](#).
- [10] W. Bingjian, L. Yapeng, L. Quan, F. Li, L. Qing, Q. Hanlin, Z. Huixin, and L. Shangqian, "Image registration algorithm based on modified GLOH descriptor for infrared images and electro-optical images," *ser. Lecture Notes in Electrical Engineering*, Springer-Verlag Berlin Heidelberg, vol. 128, 2012. [Article \(CrossRef Link\)](#).
- [11] P. Geismann and A. Knoll, "Speeding up HOG and LBP features for pedestrian detection by multiresolution techniques," *Advances in Visual Computing*, pp. 243-252, 2010. [Article \(CrossRef Link\)](#).
- [12] J. Shen, W. Yang and C. Sun, "Real-time human detection based on gentle MILBoost with variable granularity HOG-CSLBP," *Neural Computing and Applications*, vol.23, pp.1937-1948, 2013. [Article \(CrossRef Link\)](#).
- [13] M. K. Islam, F. Jahan, and J. H. Baek, "Object cataloging using heterogeneous local features for image retrieval," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 11, pp. 4534- 4555, November 2015. [Article \(CrossRef Link\)](#).
- [14] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, pp. 915-928, 2007. [Article \(CrossRef Link\)](#).
- [15] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.28, pp. 657-662, 2006. [Article \(CrossRef Link\)](#).
- [16] K. Deb, S. Imtiaz and P. Biswas, "A motion region detection and tracking method," *Smart Computing Review*, vol. 4, no. 1, pp. 79-90, 2014. [Article \(CrossRef Link\)](#).
- [17] K. Deb and A.H. Sunny, "Shadow detection and removal based on YCbCr color space," *Smart Computing Review*, vol. 4, no. 1, pp. 23-33, 2014. [Article \(CrossRef Link\)](#).

- [18] S. F. Tom, Haines and T. Xiang, "Background subtraction with dirichlet process mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 670-683, 2014. [Article \(CrossRef Link\)](#).
- [19] B. V. Vishnyakov, S. V. Sidiyakin and Y. V. Vizilter, "Diffusion background model for moving object detection," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5/W6, 2015. [Article \(CrossRef Link\)](#).
- [20] S. Nigam and A. Khare, "Integration of moment invariants and uniform local binary patterns for human activity recognition in video sequences," *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 17303-17332, December 2016. [Article \(CrossRef Link\)](#).
- [21] S. H. Anamandra and V. Chandrasekaran, "COLOR CHILD: a novel color image local descriptor for texture classification and segmentation," *Pattern Anal Applic*, Springer-Verlag London, 2015. [Article \(CrossRef Link\)](#).
- [22] T. Khanam and K. Deb, "Human and carried baggage detection & classification based on RSD-HOG in video frame," in *Proc. IEEE International Conference on Electrical and Computer Engineering*, pp. 415-418, 2016. [Article \(CrossRef Link\)](#).
- [23] W. M. S. Arnold, M. C. Dung, C. Rita, C. Simone, D. Afshin and S. Mubarak, "Visual Tracking: an experimental survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442-1468, July 2014. [Article \(CrossRef Link\)](#).
- [24] K. Khoshelham, C. Nardinocchi, E. Frontoni, A. Mancini and P. Zingaretti, "Performance evaluation of automated approaches to building detection in multi-source aerial data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, pp. 123-133, 2010. [Article \(CrossRef Link\)](#).
- [25] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen and W. Gao, "WLD: a robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705-1720, 2010. [Article \(CrossRef Link\)](#).



**Tahmina Khanam** received her Bachelor of Science (Honors) degree in Computer Science & Engineering in December 2014 from the Chittagong University of Engineering & Technology (CUET), Bangladesh. Currently, she is pursuing her Master of Science degree under the Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET), Bangladesh. Since August 2015, she has been serving as a faculty member in the Institute of Information & Communication Technology (IICT), CUET, Bangladesh. Currently, she is holding the post of Research Lecturer of the Institute. She also fondly works as an indie game developer and mobile application developer. Her research interests include computer vision, machine learning, automatic surveillance system, and pattern recognition.



**Kaushik Deb** received B.Tech and M.Tech degrees from Department of Computer Science and Engineering of Tula State University, Tula, Russia, in 1999 and 2000, respectively. He received his Ph.D. degree in Electrical Engineering and Information System from University of Ulsan, Ulsan, South Korea, in 2011. Since 2001, he has been serving as a faculty member in the Department of Computer Science & Engineering (CSE), Chittagong University of Engineering & Technology (CUET), Bangladesh. Currently, he is a Professor in Department of CSE and Dean of the faculty of Electrical & Computer Engineering (ECE), CUET, Bangladesh. He served as the Head of the Department of CSE, CUET, Bangladesh during 2012-2015. Further, he served as the Director of the Institute of Information & Communication Technology (IICT), CUET, Bangladesh during 2015-2017. His research interests include computer vision, pattern recognition, intelligent transportation systems (ITSs), and human-computer interaction.