A Fuzzy rule- based Abandoned Object Detection using Image Fusion for Intelligent Video Surveillance Systems

Preetha K G [a], Saritha S[b]

[a,b]Rajagiri School of Engineering & Technology, Rajagiri Valley P O, Kochi, India
Email:[a]preetha_kg@rajagiritech.edu.in,[b]saritha_s@rajagiritech.edu.in

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: Abandoned object/luggage is a major threat in all public scenes like hospitals, railway stations, airports and shopping malls. Abandoned luggage may contain explosive, biological warfare or smuggled goods. Abandoned object detection is the process to identify the unattended strange object within a specific time. It is also crucial to identify the person who has abandoned the luggage in the scene. Video surveillance is one of the essential techniques for automatic video analysis to extract crucial information or relevant scenes. The main objectives of this work is the automatic detection of abandoned objects and related persons in public areas like airports, railway stations, shopping malls etc. Video enhancement techniques like residual dense networks are adopted to improve the quality of the image before applying it to detect the abandoned objects and related humans. The scenario of abandoned objects and related humans are identified through distance differencing methods. Once the scene is identified, the method is capable of producing alert messages or alarms in real-time through automated means. A fuzzy rule based threat assessment module is also incorporated in this work which reduces the false alarm rate. The related person is identified through reconstruction of the face through super-resolution techniques. Experiments are found to be appreciable in terms of the metrics in video enhancement, detection, fuzzification and face super-resolution.

Keywords: Video Surveillance, Object Detection, Baggage, Fuzzy, Distance Differencing

1. Introduction

In the context of different types of attacks happening all around the world, it is seen that the lives of innocent people are affected. Considering this scenario, it is highly necessary to have an automated surveillance system that alerts real-time potential threats in the environment. Abandoned object/luggage is seen as a major threat in all public scenes like hospitals, railway stations, airports and shopping malls, as these are usually carriers of explosives, biological warfare and even smuggled goods. For this reason, it is highly necessary to detect the abandoned objects as well as the human beings who are the cause of this action. The biggest challenge faced in this area is the resolution of the low cost video surveillance system. The detection of human face is highly impossible from a low resolution video/image frames. The precision of data that is acquired from videos and images depends on the quality of the video. A poor quality video footage can lead to information loss in turn reduce the efficiency of crime investigation. The quality of video footage can affect many evidence extraction techniques such as object recognition and detection. The accuracy of most of the computer vision, image processing and machine learning tasks which is used to perform video analysis can be improved by enhancing the image appearance and quality.

The paper proposes a model to (a) automatically detect abandoned objects and (b) identify human faces who has abandoned the objects from low-resolution surveillance cameras. The automation of this model will obviously rectify the errors which arise due to human flaws resulting because of weariness and negligence. The main objectives of this work is the automatic detection of abandoned objects in public areas like airports, railway stations, shopping malls and also to identify human faces related to these abandoned objects. The main stages of the proposed system are (i) video streaming, which stream the live video and break into image instances (ii) video enhancement for improving the quality of the video (iii) object detection based on distance between the object and the associated person representing a spatiotemporal pattern (iv) fuzzified threat assessment model and (v) identification of human faces by reconstruction through super-resolution technique. The video enhancement techniques process the image and produce more enhanced image than the original input image. The fuzzified threat assessment helps to give a semantic meaning for the entire system, as visualizing through the human eye.

The rest of the paper is organized as follows. Section II review the literature in the context of this research topic. The details of the proposed method is presented in section III. Results are reported in section IV and a conclusion is drawn in section V.

2.Related Research

The increasing growth of security needs in the public area demands the need of automatic detection of abandoned objects. Abandoned object detection has become the active research area for the past few years and there are various methods for detecting abandoned object in literature. Still there is room for improvement in false alarm rate and quick response when unattended strange object is detected. The automated system capture the video from the surveillance camera installed in a particular scene. The captured video is divided into different image frames and various image processing methods are used to analyses those images. The common method in abandoned objects detection use background subtraction as a low-level preliminary step [1,2,3] to detect foreground regions or objects. A weakness of this type of methods is the increased false alarm rates caused by imperfect background subtraction due to the presence of stationary people and crowded scenes. The literature also points to the fact there are attempts to reduce false positive alarms of abandoned objects using different methods like object tracking and classification [4], edge detection [5] and generative models [6]. An inference logic based method on temporal domain is proposed in [7] which seems to be an alternative solution. The work proposed in [8] combines short-term and long-term background models to extract foreground objects. Another approach in [9] describes an algorithm for finding the stable region between image sequences. The work in [10] present a threat assessment algorithm that combines the concept of ownership in order to infer abandonment of objects. A blob tracker is used to track foreground objects based on their parameters in [11]. The concept of color representation is extended into the former model and a template matching scheme is incorporated to [12] to remove stationary objects, thus improving on the rate of false alarms. Instead of using a single camera, multiple cameras are used for detecting abandoned luggage. Auvinet et al. [13] employed two cameras for detecting abandoned objects, and the planar homography between the camera is employed to get the alarms. Kalman filter is also used to track foreground objects based on low-level features, such as color, contour, and trajectory [14]. The current state of art points to deep learning methods using convolutional neural networks [15] to find abandoned baggage, thus reducing false rates. To enhance the work, it is felt that certain steps have to be adopted in pre and post phases of this process. In the pre-processing phase, video enhancement methods have to be adopted for enhancing the video to improve the quality. Common video enhancement techniques include histogram equalization [16], contrast limited adaptive histogram equalization [17] and deep neural network techniques like convolutional neural network or residual dense network [18].

3.Proposed System

The literature review highlights the major contributions relevant to the field of abandoned object detection over the years. In the reported work, ambiguous conditions are not handled and only the systems respond to crisp scenario, thus increasing the false alarm rate. Hence fuzzy modeling is brought into the video surveillance system for better performance in the proposed model. Image fusion is also incorporated for better clarity of the image change.

The entire workflow of the proposed system is presented in Figure 1.

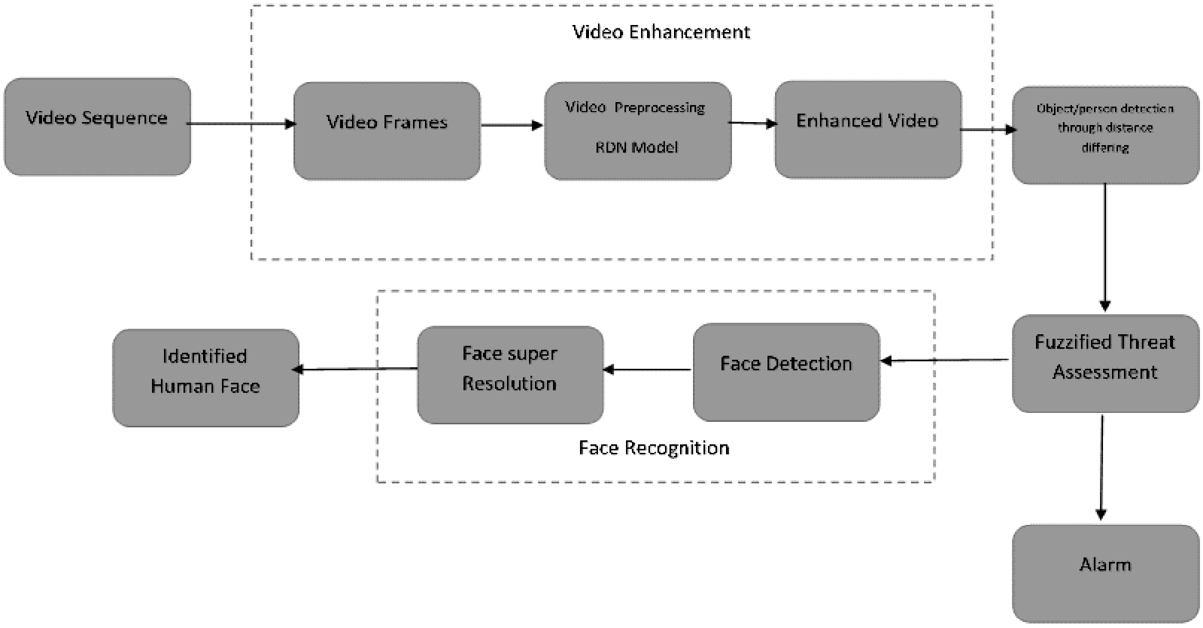

Figure 1: Abstract model of the proposed system

Video enhancement process include techniques that improve the visual quality of original data before processing. By using video enhancement methods, we can remove noise and increase the sharpness and brightness of the frames. In this work, a Residual Dense Networks (RDN) based model which is based on dense connected convolutional layers is used for the enhancement of the input video. The RDN used here extract the image features from the low quality input video frames and then fuse the relevant features from all the layers and produce enhanced video frames. Using a RDN based model, hierarchical features are extracted from the low quality image. The residual dense network contains residual dense blocks (RDB), which contains dense connected layers and local feature fusion. In this work, a pre-trained residual dense networks for video enhancement is adopted from [18].

Figure 2 brings out the major steps of difference image detection. On each image in the sequence, a standard method for background subtraction subtracts the estimated background from each video frame. The resulted foreground mask can be further processed by applying erosion and dilation filters to enhance quality of the image. Shape is a powerful signal for recognizing objects in images and segmenting images into regions corresponding to individual objects. Reliable methods are needed to detect fragments of object boundaries which is a challenging task. The shape contours are used for identifying persons and objects/bags in the images. As a starting point for contour detection, the posterior probability of contour boundary with orientation angle is estimated by measuring the difference in local image brightness, color and texture channels. The shape contour is computed through an oriented gradient signal from an intensity image. Second-order filtering techniques can be used to enhance the local maxima and smooth out peaks. The computation of space contour is motivated by the intuition that contours correspond to image discontinuities and histograms provide a robust mechanism for modeling the content of an image region. A strong oriented gradient response means a pixel is likely to lie on the boundary between two distinct regions. Thus the persons and objects are detected in all image sequences. The persons and objects detected in the images have to be validated to ensure the accuracy of detection.

Once the person and object is detected, the difference between images of adjacent time stamps are to be considered for finding the difference image. The concept of image differencing gives smaller values for unchanged pixels and bigger values for changed pixels. The difference of two images are not taken in a straightforward manner, as these difference values will not be enhanced in the region of interest. In order to enhance the difference, mean ratio operator and log ratio operators are applied on the images at two different time stamps. The main aim of any image fusion algorithm is to combine all the important visual information from multiple input images such that the resultant difference image contains more accurate and complete information than the individual source images. Appropriate image fusion technique will be adopted to perform the operation of fusion.

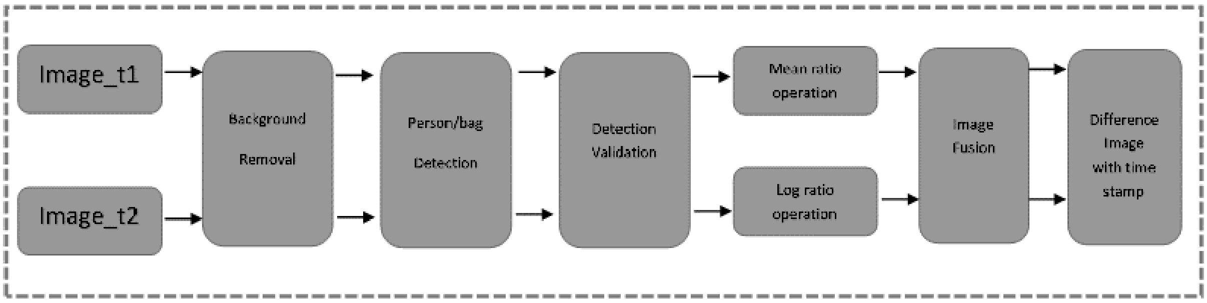

Figure 2: Steps in Difference Image Detection

Major steps of threat assessment through fuzzy techniques are portrayed in Figure 3. The difference image is then fed into the feature extractor. The feature extractor extracts features with respect to (i) Locus (position of person and bag) (ii) Behavior (variation of shape contour of the person) and (iii) Time Delay (observed time intermission between dropping object and picking). The fuzzy function of the variables X1 (Locus) X2 (Behavior) and X3 (Time) is computed as explained below.

The fuzzy membership function of the fuzzy variable X1 (Locus) is given by

$$\mu(X1) = e^{-(X1 - c_i)^2/2}, \tag{1}$$

where X1 corresponds to the crisp value of locus and c_i denotes the centre of the fuzzy function. Figure 4 represents the fuzzy membership function.

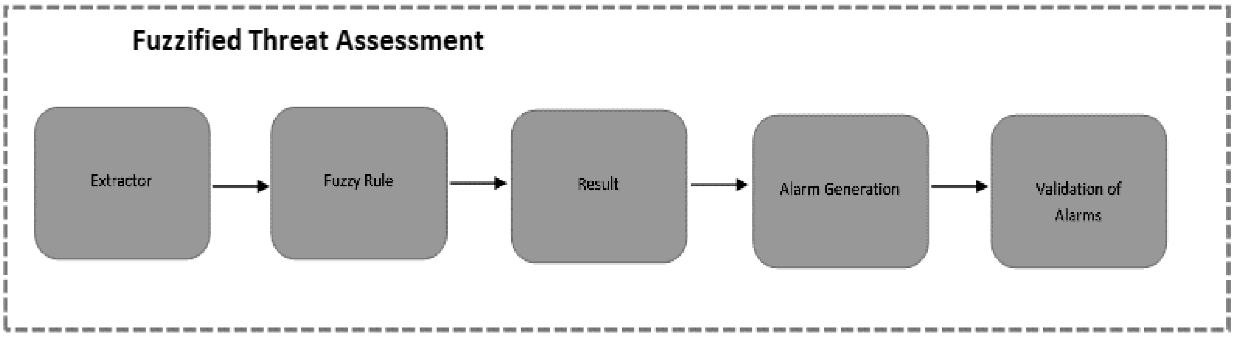

Figure 3: Threat Assessment using Fuzzy Rule Based System

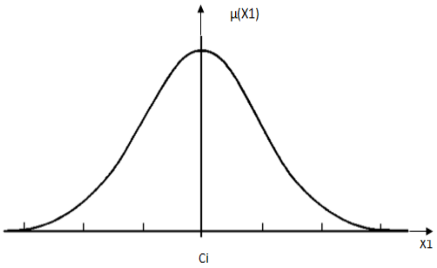

Figure 4: Fuzzy Membership function

The variable X1 is spanning over the range 0 to 1 and is considered as the support for the definition of fuzzy set. The fuzzy variable X1 assumes the five values centred on c_i as listed in Table 1 below. Similarly X2 and X3 also spanning over a range of 0 to 1 is also represented as a fuzzy function, defined over similar support (Table 2 and Table 3).

Table 1: Definition of Fuzzy Function (X1)

Fuzzy Functions of *Locus*	Membership Value (**$0<c_i<1$**)
Close	0.16
Nearby	0.32
Distant	0.48
FarAway	0.64
Infinity	0.80

Table 2: Definition ofFuzzy Function (X2)

Fuzzy Functions of *Behavior*	Membership Value (**$0<c_i<1$**)
Suspicious	0.2
NotSoSuspicious	0.4
Fair	0.6
Normal	0.8

Table 3: Definition of Fuzzy Function (X3)

Fuzzy Functions of *Time*	Membership Value (**0<c_i<1**)
Immediate	0.25
Slow	0.5
Abandon	0.75

Given the crisp values of X1 X2 and X3, corresponding to a typical instance of time, all the rules are fired and combined to produce a final inference using the crisp value Y given by

$$Y=\frac{\sum_{i=1}^{5}\sum_{j=1}^{4}\sum_{k=1}^{3}\mu_i(X1)\mu_j(X2)\mu_k(X3)c_{ijk}}{\sum_{i=1}^{5}\sum_{j=1}^{4}\sum_{k=1}^{3}\mu_i(X1)\mu_j(X2)\mu_k(X3)}$$

wherec_{ijk} corresponds to the center of the fuzzy variable indicating the inference in the rule base at the position <i,j,k> . The crisp value Y compared against an assigned threshold and choose the decision.

Sample fuzzy rules are given below.

1. *if (Locus is FarAway) and (Time is Slow) and (Behavior is NotSoSuspicious), then raise an alert*
2. *if (Locus is Infinity) and (Time is Abandon) and (Behavior is Suspicious), then raise an alarm*
3. *if (Locus is Nearby) and (Time is Immediate) and (Behavior is Normal), then disregard*
4. *if (Locus is Distant) and (Time is Slow) and (Behavior is Fair), then raise a warning*

The humans detected are then used to recognize the face. The human face detection is done using Haar cascade classifiers. The Haar features are basic features used for face detection. As the detected facial regions are of low resolution, super resolution techniques can be used to improve the resolution of the facial image. Deep learning based methods can be introduced to super resolution techniques to add the missing pixels. This work proposes a convolutional neural network layer for image up scaling.After identifying the region which include the face, the face region is cropped and given to the model which is created using convolutional layers. The output will be a high resolution facial image, which can be identified through appropriate models.

4.Results and Discussions

Experiments are run on datasets chosen from well-known sources such as PETS2006 [19] and MOTS17-01[20]. Specific videos from the dataset are used for analyzing the different scenarios outlined in the experiments.

4.1.Video Enhancement

In this module, the video frames are fed into Residual Dense Network for enhancement. In order to understand the merits of residual dense networks, it is compared with the traditional methods of enhancement like Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE). The comparison of the same is presented in Figure 5. A detailed analysis of the methods in terms of the evaluation metrics are presented in Table 4. The metrics used for evaluation are Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE). Mean squared error (MSE) calculates the average squared difference between actual and ideal pixel values. It is used to find the similarity between the images. Peak signal-to-noise ratio (PSNR) indicates the ratio of the maximum pixel intensity to the power of the distortion. The Structural similarity index (SSIM) is used for measuring the similarity between two images. It checks whether the images have similar pixel density values.

FRAMES	**PSNR (db)**			**SSIM**			**MSE**		
	HE	CLAHE	RDN	HE	CLAHE	RDN	HE	CLAHE	RDN
FRAME 1	10.1	10.16	20.9	0.30	0.23	0.73	19062	18769	1565
FRAME 2	9.8	10.47	19.9	0.27	0.22	0.81	20413	17505	1951
FRAME 3	9.9	9.73	22.8	0.33	0.21	0.78	19550	20742	1021
FRAME 4	9.6	11.14	20.5	0.32	0.28	0.73	21030	14970	1723
AVERAGE	**9.85**	**10.37**	**21.05**	**0.30**	**0.23**	**0.76**	**20013**	**17996**	**1565**

Table 4. Comparison of results of different methods of video enhancement techniques

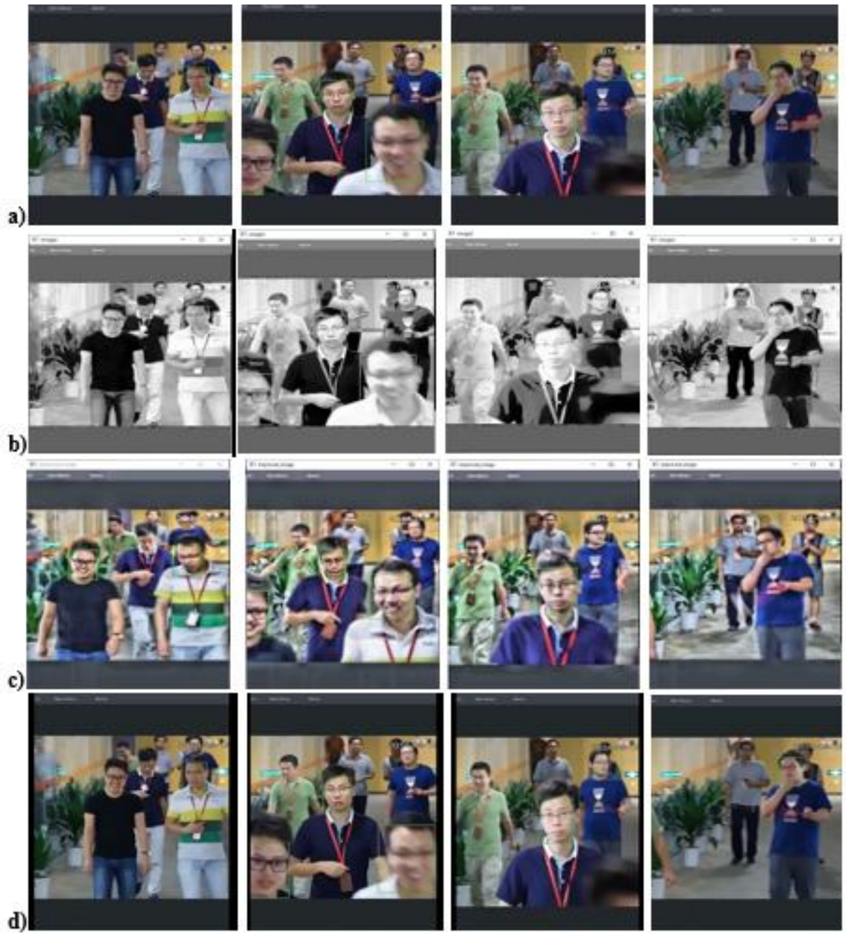

Figure 5. a) Input Frames b) Histogram Equalization (HE) c) Contrast Limited Adaptive Histogram Equalization (CLAHE) d) Residual Dense Networks (RDN)

4.2.Object/Person Detection

Shape contours and image differencing techniques are applied to detect persons and baggage. Once the objects and persons are detected, a pattern is observed for a particular time frame, which mainly focus on distance differencing method. As the distance between the baggage and the person increases, the focusing period is commenced. The threshold for the period has to be set through heuristic methods. If the distance between the person and baggage is not decreasing within the threshold value, it calls for an action and the process is taken up by the fuzzified threat assessment model. A sample example shown in Figure 6.

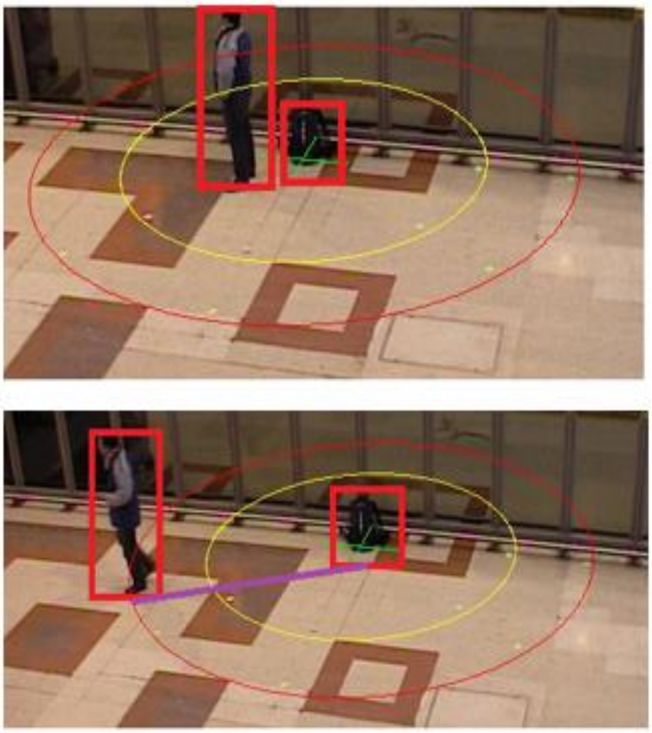

Figure 6. Object and Person detection using distance differencing method

4.3.Fuzzified Threat Assessment

Fuzzy rules are built for the different possible combinations of *locus, time and behavior*. The output is generated in the form of the options like *alert, alarm, warning and disregard*. The *locus* of the dataset chosen for experiments are set in the range of 1 m to 2 m. Experiments are run for different cases, wherein the time threshold is set in the range of 100 s – 300 s. As these two quantities are discrete ones, it is observed that they do not influence the true positive and false positive rates. The behavior of the person is chosen from shape contours and are found to be more error prone in the case of confident and fraudulent persons.

The results of the experiments are summarized as follows in the form of confusion matrix in Table 5.

	Alert	*Alarm*	*Warning*	*Disregard*
Alert	46	4	15	2
Alarm	11	52	3	4
Warning	3	12	29	5
Disregard	2	1	5	43

Table 5: Confusion Matrix

4.4.Face Reconstruction

Haar cascade classifiers are used to detect the images of faces as shown in Figure 7. The face super resolution is done to increase the resolution of the image. After cropping the facial region, it is used to apply the super resolution technique. The results after applying the super-resolution technique is given in Figure 8. The image quality metrics like PSNR, MSE and SSIM are used for evaluating the quality of the low resolution and high resolution facial images and the results are reported in Table 6. For evaluating the image quality, a facial image of each person is kept as the reference image. It is observed that the face after applying reconstruction performs appreciably well for the metrics like PSNR, MSE and SSIM.

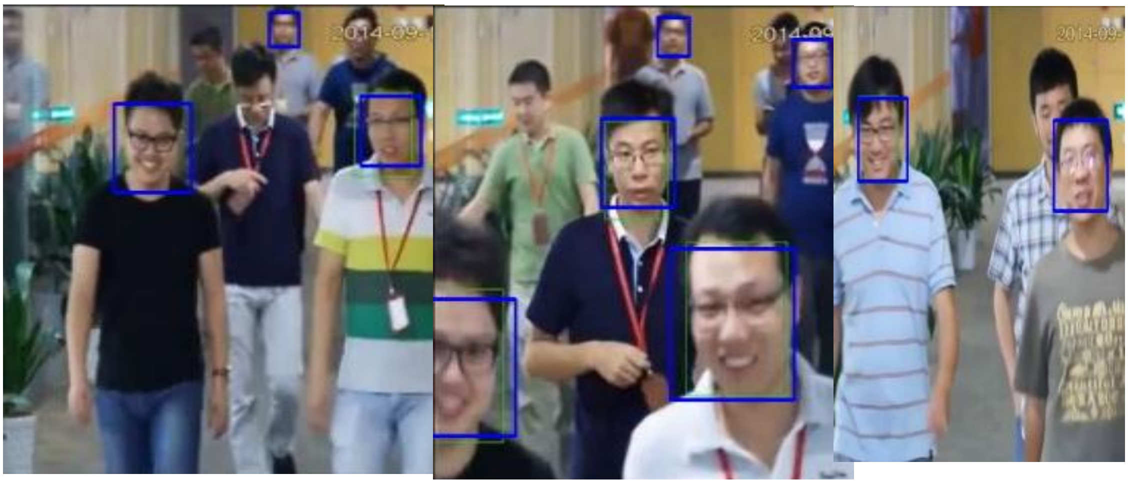

Figure 7. Face Detection

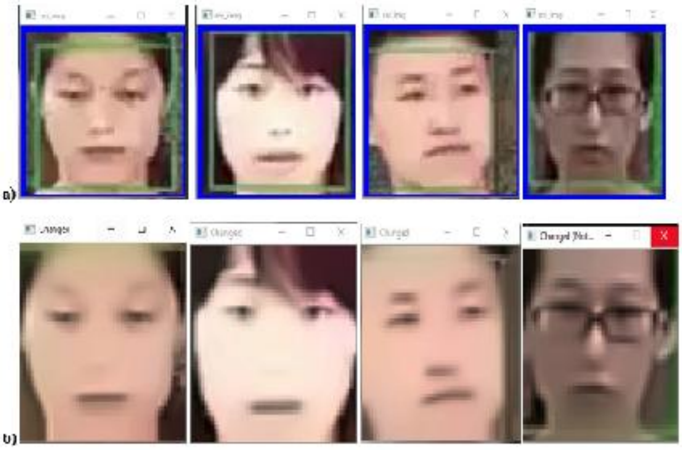

Figure 8. Face Reconstruction using super-resolution

Table 6. Image quality metrics

	PSNR (DB)		**MSE**		**SSIM**	
	Low Resolution Face	**High Resolution Face**	**Low Resolution Face**	**High Resolution Face**	**Low Resolution Face**	**High Resolution Face**
Face 1	13.50	26.87	8698	400	0.37	0.95
Face 2	14.16	27.71	7484	330	0.57	0.95
Face 3	11.58	24.66	13541	666	0.28	0.94
Face 4	12.34	23.48	11381	873	0.43	0.94
Average	**12.89**	**25.68**	**6230**	**567**	**0.41**	**0.94**

5.Conclusions

The proposed methodology is an automated system that identifies abandoned baggage and related persons in surveillance videos. It is to be noted that surveillance camera videos/images are of low quality. Hence, a video enhancement technique using residual dense network is used to enhance the images to high quality. The baggage and persons are detected through distance differencing methods. The shape contours of the person aids in supporting the fuzzy threat assessment model in this work. It is observed that the fuzzy modelling brought into the video surveillance system reduces the false alarm rate to a greater extent. By using super-resolution technique, it is also possible to reconstruct the human faces in the surveillance videos. However, further experiments are necessary to understand the results of the proposed system in low-lighting and extreme weather conditions. It is concluded that the proposed method helps to rectify the errors which arise due to human flaws resulting from weariness because of prolonged observation and consequent negligence.

References

N. T. Pham, K. Leman, J. Zhang, and I. Pek, “Two-stage unattended object detection method with proposals,” in Proceedings of ICSIP, 2017, pp. 1–4.

K. Lin, S. Chen, C. Chen, D. Lin, and Y. Hung, “Left-Luggage Detection from Finite-State-Machine Analysis in Static-Camera Videos,” in Proceedings of ICPR, 2014

P. Forczma´nski and M. Seweryn, “Surveillance Video Stream Analysis Using Adaptive Background Model and Object Recognition,” in Proceedings of ICCVG, 2010, pp. 114–1

J.-Y. Chang, H.-H. Liao, and L.-G. Chen, “Localized Detection of Abandoned Luggage,” EURASIP Journal on Advances in Signal Processing, vol. 2010, no. 1, 2010.

I. Dahi, M. Chikr el Mezouar, N. Taleb, and M. Elbahri, “An Edge-based Method for Effective Abandoned Luggage Detection in Complex Surveillance Videos,” Computer Vision and Image Understanding, vol. 158, no. C, pp. 141–151, 2017.

J. Wen, H. Gong, X. Zhang, and W. Hu, “Generative model for abandoned object detection,” in Proceedings of ICIP, 2009, pp. 853–856.

J. Ferryman, D. Hogg, J. Sochman, A. Behera, J. A. Rodriguez- Serrano, S. Worgan, L. Li, V. Leung, M. Evans, P. Cornic et al., “Robust abandoned object detection integrating wide area visual surveillance and social context,” Pattern Recognition Letters, vol. 34, no. 7, pp. 789–798, 2013.

K. Lin, S. C. Chen, C. S. Chen, D. T. Lin, and Y. P. Hung, “AbandonedObject Detection via Temporal Consistency Modeling and Back- Tracing Verification for Visual Surveillance,” IEEE Transactions on Information Forensics and Security, vol. 10, no. 7, pp. 1359–1370, 2015.

G. Szwoch, “Extraction of stable foreground image regions for unattended luggage detection,” Multimedia Tools and Applications, vol. 75, no. 2, pp. 761–786, 2016.

J. Ferryman, D. Hogg, J. Sochman, A. Behera, J. A. Rodriguez- Serrano, S. Worgan, L. Li, V. Leung, M. Evans, P. Cornic et al., “Robust abandoned object detection integrating wide area visual surveillance and social context,” Pattern Recognition Letters, vol. 34, no. 7, pp. 789–798, 2013.

F. Lv, X. Song, B. Wu, V. K. Singh, and R. Nevatia, “Left-luggage detection using Bayesian inference,” in Proc. IEEE Int. Workshop PETS, 2006, pp. 83–90.

L. Li, R. Luo, R. Ma, W. Huang, and K. Leman, “Evaluation of an IVS system for abandoned object detection on PET S 2006 datasets,” in Proc. IEEE Workshop PETS, 2006, pp. 91–98.

E. Auvinet, E. Grossmann, C. Rougier, M. Dahmane, and J. Meunier, “Left-luggage detection using homographies and simple heuristics,” in Proc. 9th IEEE Int. Workshop PETS, 2006, pp. 51–58.

J. Martínez-del-Rincón, J. E. Herrero-Jaraba, J. R. Gómez, and C. Orrite-Urunuela, “Automatic left luggage detection and tracking using multi-camera UKF,” in Proc. IEEE 9th IEEE Int. Workshop PETS, Jun. 2006, pp. 59–66.

Smeureanu, Sorina, and Radu Tudor Ionescu. "Real-Time Deep Learning Method for Abandoned Luggage Detection in Video." arXiv preprint arXiv:1803.01160 , 2018.

Wang, Qing, and Rabab K. Ward. "Fast image/video contrast enhancement based on weighted thresholded histogram equalization." IEEE transactions on Consumer Electronics 53.2 pp 757-764, 2007.

Yadav, Garima, SaurabhMaheshwari, and Anjali Agarwal. "Contrast limited adaptive histogram equalization based enhancement for real time video system." 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2014.

YulunZhang ;YapengTian ,Yu Kong ; BinengZhong ; Yun Fu “Residual Dense Network for Image Super-Resolution”, IEEE/CVF Conference on Computer Vision and Pattern Recognition , December 2018

http://www.cvg.reading.ac.uk/PETS2006/data.html

https://motchallenge.net/data/2D_MOT_2015/