Published as:

C.Beyan, A.Temizel, *"A Multimodal Approach for Individual Tracking of People and Their Belongings",* Imaging Science Journal, vol. 63, Issue 4, pp. 192-202, May 2015**.**

http://www.tandfonline.com/doi/abs/10.1179/1743131X14Y.0000000101

http://www.tandfonline.com/action/doSearch?AllField=A+multimodal+approach+for+individual+tracking+of+people+and+their+belongings

A Multimodal Approach for Individual Tracking of People and Their Belongings

Cigdem Beyan, Alptekin Temizel

Middle East Technical University

Graduate School of Informatics

Ankara, 06800, Turkey

E-mail: beyancigdem@gmail.com, atemizel@metu.edu.tr

Corresponding Author: Cigdem Beyan

ABSTRACT

In this study, a fully automatic surveillance system for indoor environments which is capable of tracking multiple objects using both visible and thermal band images is proposed. These two modalities are fused to track people and the objects they carry separately using their heat signatures and the owners of the belongings are determined. Fusion of complementary information from different modalities (for example thermal images are not affected by shadows and there is no thermal reflection or halo effect in visible images) is shown to result in better object detection performance. We use adaptive background modeling and local intensity operation for object detection and the mean-shift tracking algorithm for fully automatic tracking. Trackers are refreshed to resolve potential problems which may occur due to the changes in object's size, shape and to handle occlusion-split and to detect newly emerging objects as well as objects that leave the scene. The proposed scheme is applied to the abandoned object detection problem and the results are compared with the state of art methods. The results show that the proposed method facilitate individual tracking of objects for various applications, and provide lower false alarm rates compared to the state of art methods when applied to the abandoned object detection problem.

Keywords: Video Surveillance, Object Tracking, Abandoned Object Detection, Multimodal Tracking, Image Fusion

1. INTRODUCTION

Object tracking is an important and challenging task in video surveillance applications [1, 2]. One of the most popular techniques is the mean-shift algorithm due to its robustness, ease of implementation and computational efficiency. However, the standard mean-shift algorithm suffers from a number of problems which adversely affect tracking performance. For instance, it is not adaptable to changes in object size or shape and its performance is dependent on the selection of correct kernel size at the object initialization phase. Since the kernel shape does not always fit the object, inclusion of background information into the object model is inevitable. Additionally, the tracking might shift and even fail when the object is occluded or if the color characteristics of the background are similar to that of the object. A method using the background position on the previous and current frame to compute the target model was presented in [3]. To adapt the kernel scale and the orientation of the kernel, different approaches were proposed. For instance, the mean-shift method was combined with adaptive filtering [4]. In [5], an adaptive asymmetric kernel which provides shape and scale adaptive tracking using some heuristic was proposed. Another adaptive mean-shift tracking approach using multi-scale images was presented in [6] where a Gaussian kernel was preferred and kernel bandwidth was determined by using a log-likelihood function. To solve the problem of inclusion of background information into the object model, which especially occurs when the relocation of an object is large, a multiple object tracking method using multiple kernels was presented [7]. Multiple kernels were utilized in moving areas and background and template similarities were used to improve the convergence of the tracker. In [8], by using mean-shift for edge matching, the performance of the particle filtering tracking was improved. The iterative search property of the mean-shift algorithm was used to find the local optimal solution of a probability distribution which was then used by the particle filters.

Tracking systems utilizing only the visible band videos are successful in controlled conditions, but their performance degrades when there are instant lighting changes, shadows, smoke and unstable backgrounds. Additionally, they might not work well when the foreground object has similar color characteristics to the background. Such challenging conditions may result in false positive detections and inaccurate tracking. On the other hand, with reducing costs of thermal cameras (which has predominantly been used in military applications due to their costs until recently), thermal imagery has become more feasible and started to be utilized for civilian applications. Thermal cameras, on the contrary to the visible band cameras, are not affected by lighting changes, illumination, shadows and they can function in the absence of light sources. However, it might be difficult to detect objects in thermal images when their thermal properties are similar to the environment's thermal properties. "Halo effect" which appears around the contours of very hot or dark objects is another disadvantage of thermal video [9]. Additionally, thermal reflection is a different source of problem.

Surfaces such as wet, glass, metals reflect infrared radiation which may cause false alarms while detecting or tracking objects using only thermal images. Combining information from visible and thermal band allow overcoming the drawbacks of these two different modalities to obtain more robust systems. As an example, a study on fusion of thermal and visible band images for detecting and searching objects was proposed in [10]. A surveillance system that fuses thermal video with visible spectrum video for pedestrian detection and tracking by using rule-based decisions and heuristics was presented in [11]. Another approach used Fuzzy logic and Kalman filtering in fusion step in order to detect moving objects [12]. Moreover, [13] and [14] focused on a method that uses CCTV and thermal image fusion for object segmentation and tracking. In that methods [13, 14], objects were tracked separately in thermal and visible domain and then this information was fused by using Transferable Belief Model.

An application of object tracking is automated detection of abandoned objects [1, 15, 16, 17, 18] for public security. In such systems, the most important issue is attaining low false alarm rates while not missing the real alarms, as false alarms might render the system ineffective by causing the operators to ignore these alarms. A method using information coming from multiple cameras to generate alerts when objects move away from each other was given in [19]. Another approach using multiple cameras was presented in [15] where an alarm is generated when one of the objects is stationary and the other one gets out of the field of view. A similar method analyzing object trajectories was described in [16]. Studies that aim to detect abandoned objects in crowded environments were proposed in [20, 21]. Another abandoned object detection approach which aims supporting an operator in guarding indoor environments was presented in [17]. The proposed system is capable of tracking the owner of the object or the thief taking the object. In a different technique, objects were recognized and classified based on gradient histograms [22]. A study which handles multiple occlusions, using multi-layer detection system for abandoned object detection was proposed in [23] and an unattended object detection algorithm which includes fusion of color and shape information of static foreground objects was presented in [24]. In [25], a study that uses long-term and short-term background models to detect abandoned objects was described. Even though all of these methods are useful to detect stationary objects, false detections could still occur. For example, people standing steadily or sitting on a bench might potentially be detected as abandoned objects. In such cases, object classification or living/nonliving object discrimination should be performed to reduce the number of false alarms. Furthermore, some abandoned object detection methods generate an alarm while the owner of the belonging is still next to it which makes the system ineffective since it generates many redundant alarms. Therefore, it is necessary to develop smarter algorithms detecting and tracking the belongings and owners and associating them with each other.

Individual tracking of objects such as people and the luggage they carry is important for video surveillance applications as it would allow higher level inferences (such as abandoning of an object by a person) and make timely detection of potential threats possible. Additionally, by using the trajectories of people and their belongings, interactions between them could be deduced. However, this is a challenging problem and in the literature, people and objects they carry are commonly tracked as a single object. Methods for abandoned object detection and abandoned object-owner matching were presented in [17, 18]. However, neither of these studies tracked people and their belongings separately. In [17], the owner of an object was determined when the object and its owner split. However, this might result in false owner association and cause false alarm in the case of occlusion of owner or abandoned object. Moreover, two persons who are entering the scene together might cause a false alarm when they split since one of the persons is likely to be detected as an abandoned object. It might also be difficult to discriminate the object by using its size since the luggage and a child's or a sitting person's observed sizes in the image could be similar. Method proposed in [18] firstly detected the abandoned object and then searched through the history to find the owner of the abandoned object. Later, it associated the people with abandoned object by calculating the overlap of the bounding boxes of the owner and abandoned objects. However, this method requires memory to keep locations of each tracked object for all frames and also does not provide a mechanism for differentiating people and belongings. For instance, given that two people walk in proximity and one of them sit after a while. Such a system [18], will detect the person who sits as an abandoned object while the other person will be assigned as the owner of that false abandoned object.

In this study, we propose a fully automatic multiple object tracking algorithm using thermal imagery in addition to the visible band imagery for indoor applications. Different to the other studies, visible and thermal domain tracking information is fused to track people and the objects they carry separately using their heat signatures. Utilizing the trajectories of these objects obtained by individual tracking, *i)* interactions between them are deduced, *ii)* owners of the objects are associated and *iii)* abandoned objects are detected to generate alarms. Additionally, better detection performance is also achieved compared to using a single modality: thermal reflection and halo effect which adversely affect tracking are eliminated by the complementing visible band data and thermal data helps increasing tracking performance when visible band contrast is low due to object color, shadows or insufficient illumination.

The paper is organized as follows: Section 2 includes the detailed information about the proposed method. Section 3 contains experimental, quantitative results and information about the experiment dataset. Lastly, the concluding remarks are given in Section 4.

2. PROPOSED METHOD

The main steps of the proposed method are illustrated in Figure 1. As shown in this figure, firstly background subtraction is applied to the visible band image. Then, connected component analysis is utilized to remove the noise. On the other hand, local intensity operation is applied to the thermal image and the result of this operation is post-processed to complete and close possible holes which might be formed after local intensity operation. Object discrimination step is the fusion step which uses both modalities. We apply fusion at this stage as it was reported that fusion after object detection approach is reported to be the best performing scheme [18]. After this step, a rule based method and connected component analysis are used to extract objects and classify them as people or belongings. Finally, each object (people and/or belonging) is tracked using the improved, adaptive mean-shift tracking algorithm. While tracking objects, people and belongings are also associated with each other and owner/carried object relation is set for tracked objects. Abandoned objects can be detected by using these relations and tracking the objects separately. These steps are described in more detail below.

2.1. Background Subtraction and Noise Removal

Improved Adaptive Gaussian model proposed in [26] is a technique that was reported to produce reliable background information while being computationally not very complex. This is a pixel-based method and each pixel is defined as a mixture of Gaussians with *M* components as follows:

$$\hat{p}\left(\vec{x}\middle|X_{T,}BG+FG\right)=\sum_{m=1}^{M}\hat{\pi}_m N(\vec{x};\widehat{\vec{\mu}_m},\widehat{\sigma_m^2}I) \qquad (1)$$

where $x^{(t)}$ is the value of pixel at time t, $X_T=\{x^{(t)},\ldots, x^{(t-T)}\}$ is the training set at time *t* while *T* is the time period, *BG* is the background, *FG* is the foreground, μ_1, μ_2 , …., μ_M and σ_1 , σ_2 ,…., σ_M are the estimates of mean and variance for the Gaussian components respectively. $\pi_1, \pi_2, \ldots, \pi_M$ are the weight values that are nonnegative and whose summation is equal to 1. The parameters of the model should be updated with new samples to adapt to the changes in the background [26].

Although this method is based on mixture of Gaussians, it is more adaptive and robust as it could automatically select the proper number of components per pixel [26].

After background subtraction, we apply connected component analysis to detect and remove the noise. To eliminate the noise, the bounding box of the object, the

number of pixels that each connected component has and the area of the object's bounding box are found. Then, density of each object is calculated by using Eq. 2.

$$D = N/A_{rect} \tag{2}$$

where D is the density of object, N is the total number of pixels of the object and A_{rect} is the area of the bounding rectangle. A connected component is classified as noise and removed from the image if its density is smaller than the density threshold and the number of pixels that belongs to this object is smaller than the maximum number of pixel threshold. An example result of background subtraction and noise removal step is shown in Figure 2.

2.2. Local Intensity Operation for Thermal Images

Thermal domain images are constructed from thermal energy emitted by objects and in white-hot setting of thermal camera, pixels of living objects appear brighter than most background objects. The study in [27] uses local intensity operation (LIO) for defect detection in thermal images. We adapt this operator, which brightens the bright pixels and darkens the dark pixels, for our purpose of segmenting pixels potentially belonging to people as follows:

For a pixel $I(x,y)$ in thermal image, denoted with z_0, Z value is calculated as the product of the pixel and its eight neighbors (z_1 to z_8):

$$Z = \prod_{k=0}^{8} Z_k \tag{3}$$

Then a new image is created by calculating Z for each pixel in thermal image as in Eq. (4).

$$g(x, y) = Z \tag{4}$$

where $g(x, y)$ is the pixel value at (x, y) of new image.

After that, these image pixels are normalized to gray-scale range. Although, this operation increases the brightness of bright pixels and the darkness of the dark pixels to get better result, we segmented this new image by using Mean Absolute Thresholding (MAT) Eq. (5).

$$T = \text{round}\left[\frac{I_{max} - I_{min}}{2}\right] \tag{5}$$

where T is the threshold value, I_{max} is the maximum pixel value, I_{min} is the minimum pixel value.

Examples of algorithm's result are shown in Figure 3.

It has to be noted that, besides the living objects, hot objects such as heating systems, radiators, television screens (commonly used in airports as information screens) or any object which is hotter than the background are also captured brighter than the other objects and segmented as a result of this process. However, as will be explained in later sections, since such objects belong to the background in the visible image, the proposed method does not discriminate these objects as people and hence, false alarms due to stationary hot objects are prevented. Additionally, in order to overcome potential problems which may occur due to the moving pictures on television screen, a mask which excludes the area of television screen in the image could be used.

2.3. Post Processing

The algorithm explained in Section 2.2 results in segmentation of the object. However the segmentation might be imperfect and there might be gaps on the segmented object especially due to the clothing preventing thermal radiation. To have better segmentation of the objects, first, objects in binary images (result of local intensity operation), Figure 3(b) are completed by hole-filling. Then, these binary objects are closed.

2.4. Object Discrimination

Automated individual tracking of objects by only using visible band is a challenging problem since people and the objects they carry are segmented as a single object as shown in Figure 4(c). In this study, we propose a solution for this problem using the heat signature obtained from the thermal band.

Object discrimination step is the fusion phase where both the thermal and the visible band images are used and it is the main step for individual tracking of objects such as people and their belongings.

In this step, the objects obtained after foreground detection and noise removal (a binary image) in visible data (Section 2.1) and the objects obtained using the local intensity operation and post processing (a binary image) on the thermal data (Section 2.2 and 2.3) are utilized.

Using the rule given in (Eq. 6) and connected component analysis, objects are extracted and classified as people or belongings.

$$F(x,y) = \begin{cases} people, & R_V(x,y) \neq 0 \ \wedge \ R_T(x,y) \neq 0 \\ belongings, & R_V(x,y) \neq 0 \ \wedge \ R_T(x,y) = 0 \end{cases} \quad (6)$$

where $F(x, y)$ is the fusion result, $R_V(x, y)$ is the pixel value after background subtraction and noise removal in visible data and $R_T(x, y)$ is the pixel value after local intensity operation and post processing in thermal domain. This rule prevents thermal reflection and hot objects such as heating systems being classified as people.

After this operation, discrimination errors may be observed, especially around the people due to inaccuracies in the registration phase of thermal and visible band images. To handle these errors, same method which is presented in Section 2.1 is applied to remove noise and errors are eliminated. Example results of this step are shown in Figure 4 where the person is shown in red while his belonging is shown in green.

2.5. Improved, Adaptive Mean-Shift Tracking Algorithm

The mean-shift tracking method [28] is an optimization algorithm based on object representation. It uses a nonparametric kernel and iterates until the goal is attained. It aims to find an object in the next image frame which is most similar to the initialized object (object model) in the current frame. It uses the histogram of the object model and histogram of the candidate object and maximizes the likelihood using similarity between the two histograms.

In this study, the adaptive mean-shift tracking algorithm proposed in [29] was adapted to track people and their belongings individually in a multi modal setting. In this context, for initialization of the objects' bounding boxes, results of the object discrimination step (Section 2.4) are used. Then for each belonging or person, a separate tracker is defined. Additionally, the bounding box of the tracked object is used as a mask to decrease the search area of the mean-shift tracker. This increases tracking accuracy and performance of the system. Due to the reduced search area in the frame, the required number of iterations to find the new position of the object is decreased.

Although using the information coming from object discrimination phase in the tracker initialization step has advantages, it is not sufficient to make the system fully automatic since it still needs to detect the new objects entering the scene or the objects leaving the scene. To solve this and to have a system which is adaptable to change in object's size and shape and to handle inclusion of background information, we reinitialize the mean-shift trackers by using result of object discrimination step in regular time intervals. This update mechanism is summarized in Figure 5 [29].

To handle the changes in size or shape we update mean-shift trackers every second (when frame number % *fps* = 0, where fps stands for frames per second). To detect new objects as well as objects that leave the scene, numbers of objects in the adjacent frames are compared and if those numbers are not equal then mean-shift trackers are updated. To handle occlusion and split and to detect newly emerging objects; the location of bounding box of each objects are compared. If an intersection exits then mean-shift trackers are refreshed to handle the inclusion of front object's color. Separate mean-shift trackers are initialized for each people and belongings and the same algorithm is used independent of the object type.

2.6. Association of People and Their Belongings

While tracking objects, people and their belongings are associated and ownership relation is determined using a closeness criterion. The Euclidian distances between each person and the objects are calculated and the object is assigned to the nearest person. While determining the ownership, it is assumed that the object is not handover to another person. Therefore, once belonging is appointed to a person, it is not appointed to another person later. Since a wrong association is strongly possible during occlusions, association is delayed until the split, If a belonging is associated with a person while that person is occluded by another belonging, it is assigned to the person after split occurs. If these objects form a new object by merging, then in the next update, belonging is associated with the merged object. Example results of this association step are given in Figure 6. In this figure, bounding boxes of belongings are shown in red and bounding boxes of people are shown in green. To denote the association of a belonging with a person, belonging is indexed with a notation ([*Owner Index*].[*Object Index for The Owner*]). For example in Figure 6, 1.1 is the object belonging to person 1 and 2.1 is the object belonging to person 2.

2.7. Abandoned Object Detection

Abandoned object detection is an example application that can be used to demonstrate the proposed method. As the proposed system allows detection and tracking of people and their belongings separately, it allows the owner of the unattended luggage to be found. For this application, association of people and their belongings is useful as it allows finding the owner of an unattended luggage.

A belonging is detected as an abandoned object when its owner leaves the field of view and the alarm is set off after a predefined *N* number of frames. The alarm is disarmed immediately when the unattended object is removed. To prevent false alarms in the case of merging of objects, the object's owner is checked whether it is occluded and formed a new object or not (Figure 7).

3. EXPERIMENTAL RESULTS

3.1. Dataset

The proposed method was tested for 14 different scenarios containing various sizes and colors of bags when multiple occlusions exist. To prove the proposed method is not affected from hot objects in the scene, such as heating system, radiators, it was also tested in environments using such objects. As there were no public datasets for abandoned object detection and only a few limited videos for

object tracking having both modalities, we captured our own dataset for various scenarios. Table 1 shows the details of the each image sequence in the dataset. Video sequences are publicly available at: http://ii.metu.edu.tr/node/487

Thermal videos were captured using an OPGAL EYE-R640 un-cooled infrared camera and visible band videos were captured using a Sony HDR-HC1 camera. Both modalities were captured at 320x240 resolutions and 25 frames/second.

3.2. Results

Firstly, images captured from thermal and visible cameras were registered. Both thermal and visible band cameras should be adjusted properly to capture a similar field of view. However, it is not practically possible to capture exactly the same field of view (FOV) for both thermal and visible band cameras since these cameras have different parameters (such as different sensor types and lenses). Therefore, a crop operation was performed for both thermal and visible band frames to have similar FOV for both thermal and visible images. Then, homography was performed manually by selecting reference points in both thermal and visible domain for the image registration and calculating the homography matrix using Eq. (7) and (8).

$$V_{ref} = H \times T_{ref} \tag{7}$$

$$\text{H}=V_{ref} \times T_{ref}^{-1} \tag{8}$$

where V_{ref} is the reference point matrix for visible domain, T_{ref} is the reference point matrix for thermal domain, and H is the homography matrix for registration. In this study, 20 reference points were selected for each dataset. Once the capture and homography parameters are obtained, these parameters can be used as long as the camera positions are not changed. Therefore, in real life systems, when the camera positions are fixed and registration parameters are set, the proposed method can work without requiring any user intervention.

Example results of the proposed method are given in Figures 7-8.

As it is seen from the example results (Figures 6- 8), the proposed method successfully detects occlusions and split and finds the correspondence after merging of objects. Multiple occlusions are also handled. Mean-shift tracking can adapt to the changes in objects' size and shape as the trackers are refreshed. By comparing the numbers of objects in consecutive frames, new objects or objects which are leaving the scene are immediately detected. Use of the update mechanism makes the tracking system fully automatic.

Tracking performance of the proposed method was evaluated in terms of recall, precision and accuracy and compared with the standard mean-shift tracking [28] and when only visible band is used [29]. While testing with standard mean shift tracking, the minimum bounding box covering the whole object was manually selected when an object enters the scene.

For calculating the recall and precision values; manually extracted ground truth information (bounding boxes of the objects) was used. Recall is calculated as the ratio of true positives (*TP*) to the sum of *TP* and false negatives (*FN*) while precision is calculated as the ratio of *TP* to sum of *TP* and false positives (*FP*). While determining these metrics all *TP*, *FN* and *FP* values were calculated based on pixel count where *TP* is the total number of pixels where ground truth and the proposed method agree on these pixels belonging to an object. *FN* is the total number of pixels that ground truth denotes the pixel a part of the object while the tracking system cannot detect these pixels as part of an object. *FP* is defined as the number of pixels that the tracking system finds as an object while ground truth does not agree. On the other hand, while calculating the tracking accuracy the Euclidean distance between the center of mass of the estimated objection position and the center of mass of the ground truth are used.

To illustrate the tracking accuracy Set 14 starting from frame number 50 to 165 was used. The calculated Euclidean distances (in pixels) between the estimated object position and the ground truth against the frame number for each method are given in Figure 9.

In Figure 9, accuracies belong to the tracking of both a person and its belonging using the proposed method, the standard mean shift tracking algorithm and the proposed method using only visible band data are shown. As the standard mean-shift tracking algorithm is not able to detect objects automatically, any person entering the scene is determined manually. It has to be noted that for the standard mean-shift tracking algorithm, tracking results for the belonging is not available as it cannot be detected.

The proposed method has better accuracy for tracking the person when both visible and thermal images are used (shown as Proposed(Fusion)-Person) compared to when only visible band images are used (Proposed(Visible)-Person) and the standard mean-shift tracking algorithm (StandardMeanShift-Person). On the other hand the accuracy for tracking the belongings is better when only visible band is used (Proposed(Visible)- Belonging) compared to the proposed method when both bands are used (Proposed(Fusion)-Belonging). However, as the thermal clues are not available in this case, the belonging can only be detected after it is left by the person and tracking cannot be carried out while it is carried by the person. As a consequence of this, tracking results are only available after frame number 86 onwards and the belonging cannot be tracked until this frame.

Besides, the object cannot be classified as a belonging or person due to unavailability of thermal information.

In Table 2, the average Euclidean errors, precision and recall values while tracking a person and the belongings are given for each method. The best results are emphasized in boldface and the cases not available are shown as NA. Calculating the performance measures for standard mean-shift tracking algorithm for belongings is not possible since standard mean-shift tracking algorithm is not able to detect the left objects. Even though the Euclidean error and precision results for tracking the belonging with the proposed method (visible band only) seems to be higher, these metrics reflect partial results as explained above.

As seen from Table 2, the most accurate results were obtained using the proposed method. Additionally, the belonging and the person could not be classified and tracked individually by only visible-band tracking. On the other hand, the results of standard mean-shift tracking algorithm was much lower than the other methods, since it is not adaptable to change in size or the shape of the object which is being tracked.

The method (applied to abandoned object detection problem) is also compared with an abandoned object detection algorithm [30] which does not involve tracking. In this problem, the aim is to prevent false alarms due to stationary living objects and raising alarms only for abandoned nonliving objects. The false and true detection performances (using datasets listed in Table 1) of the proposed method and method in [30] are given in Table 3.

In Table 3, the best results for each scenario are emphasized in boldface and no alarm cases are shown as NA as true detection does not apply to these scenarios.

As seen from this table, neither of the methods missed the true alarm cases. On the other hand, no false alarm was given by the proposed method while method [30] caused false alarms for Sets 4, 5, 6 and 11 since it generated an alarm although the owner of belonging stays next to the belonging. Additionally, it lost the abandoned object when the object was stationary for a while since these objects become progressively a part of the background. On the other hand, the proposed method prevents false alarms in either of these cases. In the former case, it keeps track of the interactions between the person and his/her belonging and does not raise a false alarm. In the later, as the tracker does not depend on background subtraction, it continues tracking the object even when the abandoned object is stationary.

The most computationally complex parts of the proposed system are the background subtraction, connected component analysis and mean shift tracking algorithm steps which are also common steps in the methods in the literature. In

addition to these steps, the proposed system requires calculation of the homography matrix, fusion and subsequent post-processing. Calculation of the homography matrix is done only once per camera installation and is not a run-time component. Local intensity operation is only applied to thermal images and requires 8 multiplication operations followed by a summation operation for each pixel. Fusion, subsequent post-processing and association of owners and belongings steps are only applied to detected object regions and as such, have low computational complexity. As a result it can be said that the proposed system adds minor computational cost compared to common object detection and tracking based applications.

4. CONCLUSIONS

In this study, we proposed a fully automatic framework for individual tracking of multiple objects. In addition to visible band, thermal band images were also used and these two modalities were fused after object detection, allowing tracking of people and the objects they carry separately. By using the trajectories of these objects, owners of the belongings can be determined and abandoned objects can immediately be detected when left unattended by their owners. Better detection and tracking performances were also achieved as fusion of the information from different modalities eliminates the shortcomings when only one of these modalities were used (such as shadows in visible band and thermal reflection in thermal band). The information coming from object discrimination step was used to make the system fully automatic as new objects entering the field of view and objects that are leaving the scene could be detected immediately. The results showed that the method is applicable to real life scenarios. Additionally, it performed favorably with other methods in terms of false alarm rates.

It has to be noted that the proposed framework allows integrating other methods to segment the objects, fuse different modalities and track the objects. For instance, in some scenarios certain body parts of the people (especially hair and feet; hair in general is cooler than body temperature and shoes prevent body heat to be radiated) is removed as a result of the proposed fusion step, since it usually could only be detected in visible band while could not be detected in thermal band and eliminated as noise. Therefore, if tracking accuracy is important, then developing a fusion technique considering such scenarios and integrating into the fusion step of the framework could increase the system performance.

Besides detection of abandoned objects, the proposed scheme could also be used in scenarios where individual tracking of people and their belongings is beneficial. For example, the scheme could be used to count people with or without bags entering/leaving shops or shopping malls. It could also be used to detect someone picking up or stealing an object by adding proper rules following the proposed tracking system.

REFERENCES:

[1] Uke N. J., Thool R. C., Motion tracking system in video based on extensive feature set, The Imaging Science Journal, vol. 62, no. 2, pp. 63-72, (2014).

[2] Adeli-Mosabbeb E., Fathy M., Zargari F., Model-based human gait tracking, 3D reconstruction and recognition in uncalibrated monocular video, The Imaging Science Journal, vol. 60, no. 1, pp. 9 – 28, (2012).

[3] Boonsin M., Wettayaprasit W., and Preechaveerakul L., Improving of Mean Shift Tracking Algorithm Using Adaptive Candidate Model, Proc. of ECTI-CON, pp. 894-98, (2010).

[4] Qifeng Q., Zhang D., and Peng Y., An adaptive selection of the scale and orientation in kernel based tracking, Proc. of IEEE Conference on Signal-Image Technologies and Internet-Based Systems, pp. 659–664, (2007).

[5] Quast K. and Kaup A., Scale and shape adaptive mean shift object tracking in video sequences, Proc. of 17th European Signal Processing Conference (EUSIPCO), pp. 1513–1517 (2009).

[6] Jiang Z., Li S., and Gao D., An Adaptive Mean Shift Tracking Method Using Multiscale Images, Proc. of International Conference on Wavelet Analysis and Pattern Recognition, (2007).

[7] Porikli F. and Tuzel O., Multi-kernel object tracking. Proc. of IEEE International Conference on Multimedia and Expo, (2005).

[8] Liu Y. M., Zhou S. B., A self-adaptive edge matching method based on mean shift and its application in video tracking, The Imaging Science Journal, vol. 62, no. 4, pp. 206-216, (2014).

[9] Kumar P., Mittal A., and Kumar P., Study of Robust and Intelligent Surveillance in Visible and Multimodal Framework, Informatica, vol. 32, pp. 63-77, (2008).

[10] Fay D. A., Waxman A. M., Aguilar M., Ireland D. B., Racamato J. P., Ross W. D., Streilein W. W., and Braun M. I., Fusion of multi-sensor imagery for night vision: color visualization, target learning and search, Proc. of International Conference on Information Fusion, pp. 215-219 (2000).

[11] Torresan H., Turgeon B., Ibarra-Castanedo C., Hebert P. and Maldague X., Advanced Surveillance Systems: Combining Video and Thermal Imagery for Pedestrian Detection, SPIE Thermosense XXVI, pp.506–15 (2004).

[12] Kumar P., Mittal A. and Kumar P., Fusion of Thermal Infrared and Visible Spectrum Video for Robust Surveillance, Proc. of 5th Indian Conference on

Computer Vision, Graphics and Image Processing (ICVGIP), vol. 4338, pp.528–539 (2006).

[13] Conaire C. Ó, Connor N. O, Cooke E., Smeaton A., Multispectral Object Segmentation and Retrieval in Surveillance Video, Proc. of IEEE International Conference on Image Processing (ICIP), (2006).

[14] Conaire C. O., Connor N. O., Cooke E., Smeaton A, Comparison of Fusion Methods for Thermo-Visual Surveillance Tracking, Proc. of International Conference on Information Fusion, (2006).

[15] Martinez-del-Rincon J., Herrero-Jaraba J. E., Gomez J. R., and Orrite-Urunuela C., Automatic left luggage detection and tracking using multi-camera UFK, Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 59–66 (2006).

[16] Krahnstoever N., Tu P., Sebastian T., Perera A., and Collins R., Multi-view detection and tracking of travelers and luggage in mass transit environments, Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 67–74, (2006).

[17] Ferrando S., Gera G., Massa M., and Regazzoni C., A new method for real time abandoned object detection and owner tracking, Proc. of IEEE International Conference on Image Processing (ICIP), pp. 3329–3332 (2006).

[18] Denman S., Lamb T., Fookes C., Chandran V. and Sridharan S., Multi-spectral fusion for surveillance systems, Computers and Electrical Engineering, vol. 36, no. 4, pp.643-663, (2010).

[19] Auvinet E., Grossmann E., Rougier C., Dahmane M., and Meunier J., Left-luggage detection using homographies and simple heuristics, Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 51–58 (2006).

[20] Smith K**.**, Quelha P., and Gatica-Perez D., Detecting abandoned luggage items in a public space, Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 75–82 (2006).

[21] Guler S. and Farrow M. K., Abandoned object detection in crowded places, Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 99–106, (2006).

[22] Miezianko R. and Pokrajac D., Detecting and recognizing abandoned objects in crowded environments, Proc. of Computer Vision Systems, pp. 241–250 (2008).

[23] Denman S., Sridharan S., and Chandran V., Abandoned object detection using multi-layer motion detection, Proc. of International Conference on Signal Processing and Communication Systems, (2007).

[24] SanMiguel J. C., Martinez J. M., Robust unattended and stolen object detection by fusing simple algorithms, Proc. of IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS), pp. 18-25, (2008).

[25] Porikli F., Ivanov Y., Haga T., Robust abandoned object detection using dual foregrounds, EURASIP Journal on Advances in Signal Processing, (2008).

[26] Zivkovic Z., Improved adaptive Gaussian mixture model for background subtraction, Proc. of International Conference on Pattern Recognition (ICPR), pp. 28-33, (2004).

[27] Heriansyah R. and Abu-Bakar S.A.R., Defect detection in thermal image for nondestructive evaluation of petrochemical equipments, NDT & E International, vol. 42, no. 8, pp. 729-740, (2009).

[28] Comaniciu D., Ramesh V., and Meer P., Kernel-based object tracking, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 564–77, (2003).

[29] Beyan C., Temizel A., Adaptive Mean-Shift for Automated Multi Object Tracking, IET Computer Vision, vol. 6, no. 1, pp. 1-12, (2012).

[30] Beyan C., Yigit A., and Temizel A., Fusion of Thermal and Visible Band Video for Abandoned Object Detection, SPIE Journal of Electronic Imaging, vol. 20, 033001 (2011); doi:10.1117/1.3602204.

Figure Captions

Figure 1: Block diagram of the proposed system.

Figure 2: (a) Visible band image (b) Result of background subtraction (c) Result of noise removal.

Figure 3: (a) Two unprocessed thermal images, (b) Segmentation results for these images.

Figure 4: Results for two set of images (a) Visible band images, (b) Corresponding thermal images, (c) Segmentation result when only visible band is used, (d) Object discrimination results with error, (e) Segmented people, (f) Segmented belongings.

Figure 5: Improved, Adaptive Mean-shift Tracking.

Figure 6: Association of objects with their owners (Images taken from Set 12).

Figure 7: An example result of the proposed method for Set 5. Person 3 leaves her backpack (object 3.1) on the floor. After it is detected as an abandoned item, even though temporary occlusions occur due to moving people (5 and 6), those do not cause false tracking or false alarm. The alarm is raised (frame #556) after the person owning the backpack leaves.

Figure 8: An example result of the proposed method for Set 6. No false alarm is given although handbag (object 1.1) is stationary for 913 frames as its owner (person 1.1) stays next to it. There are multiple occlusions due to other people during this period.

Figure 9: Illustration of tracking accuracy of the given methods based on Euclidean distances (in pixels) between the estimated object position and the ground truth against the number of frame number.

Table Captions

Table 1: Details of the videos used in the evaluation. All video sequences have a resolution of 320x240 pixels and captured at 25 fps.

Table 2: Performance analysis for the proposed method, proposed method while only visible band is used and the standard mean-shift tracking. **The object is only detected at frame 86 and the performance metrics cannot be calculated from the beginning of the scene until frame 86.*

Table 3: Performance comparison of the proposed method applied to abandoned object detection and the method proposed in [30].

Figures

Figure 1

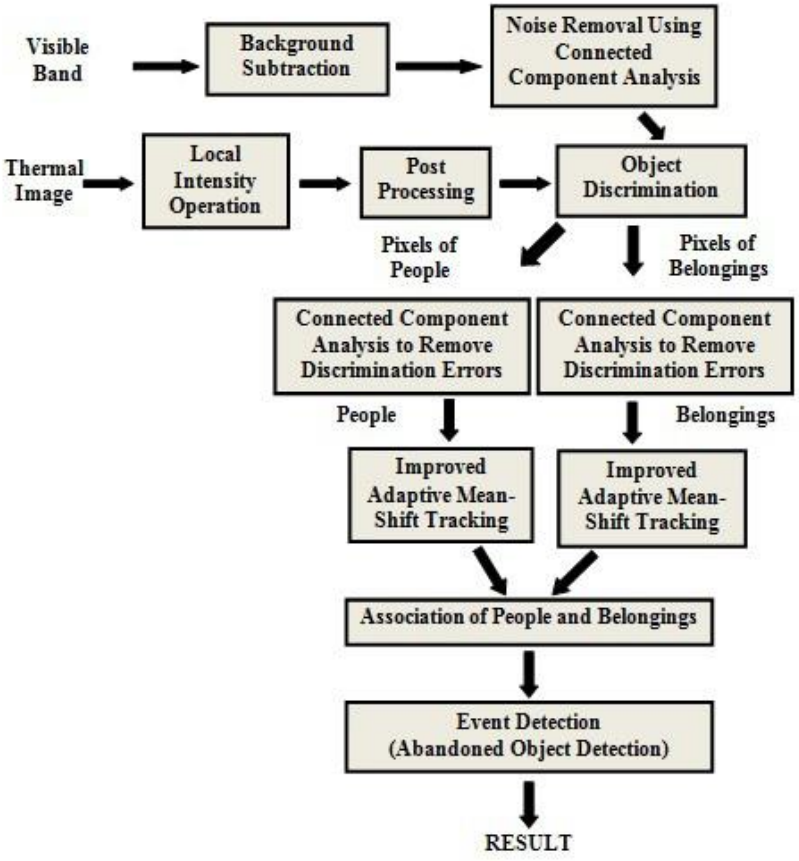


Figure 2

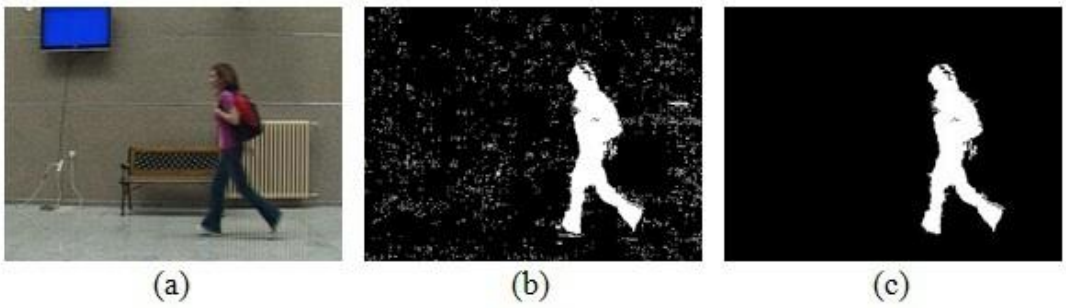


Figure 3

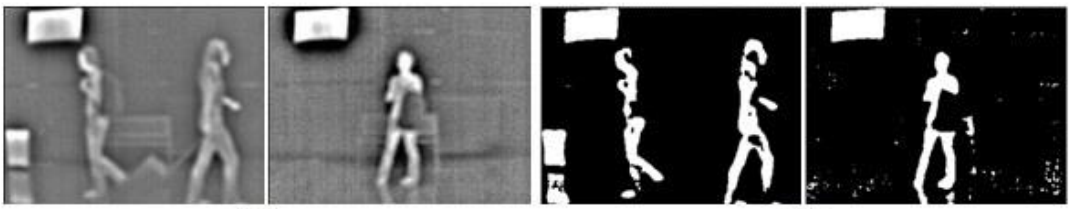


Figure 4

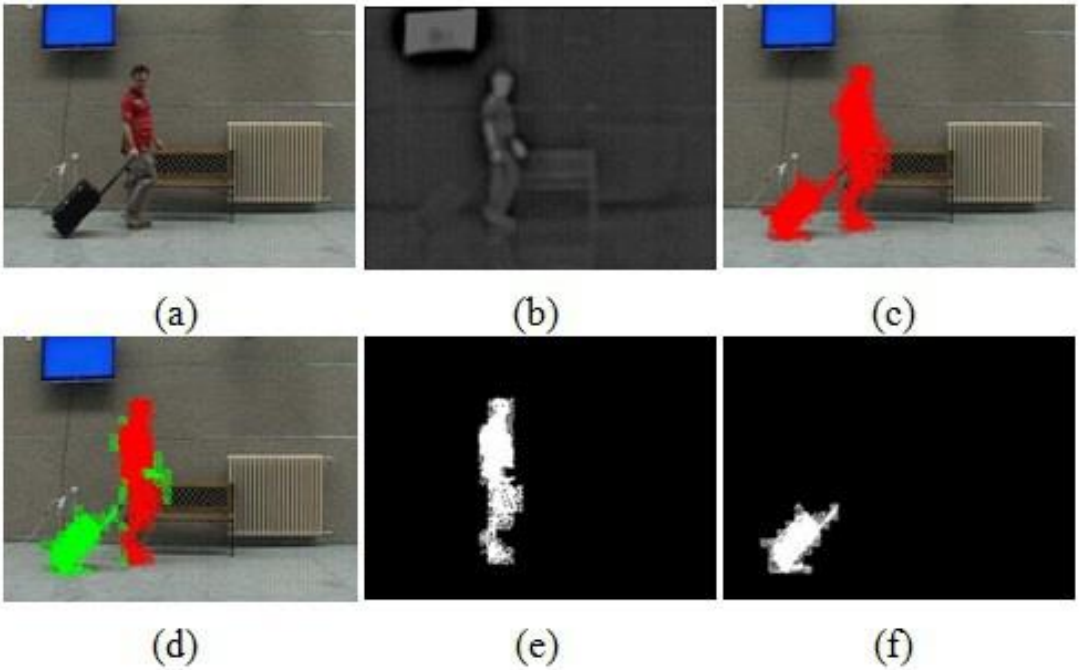


Figure 5

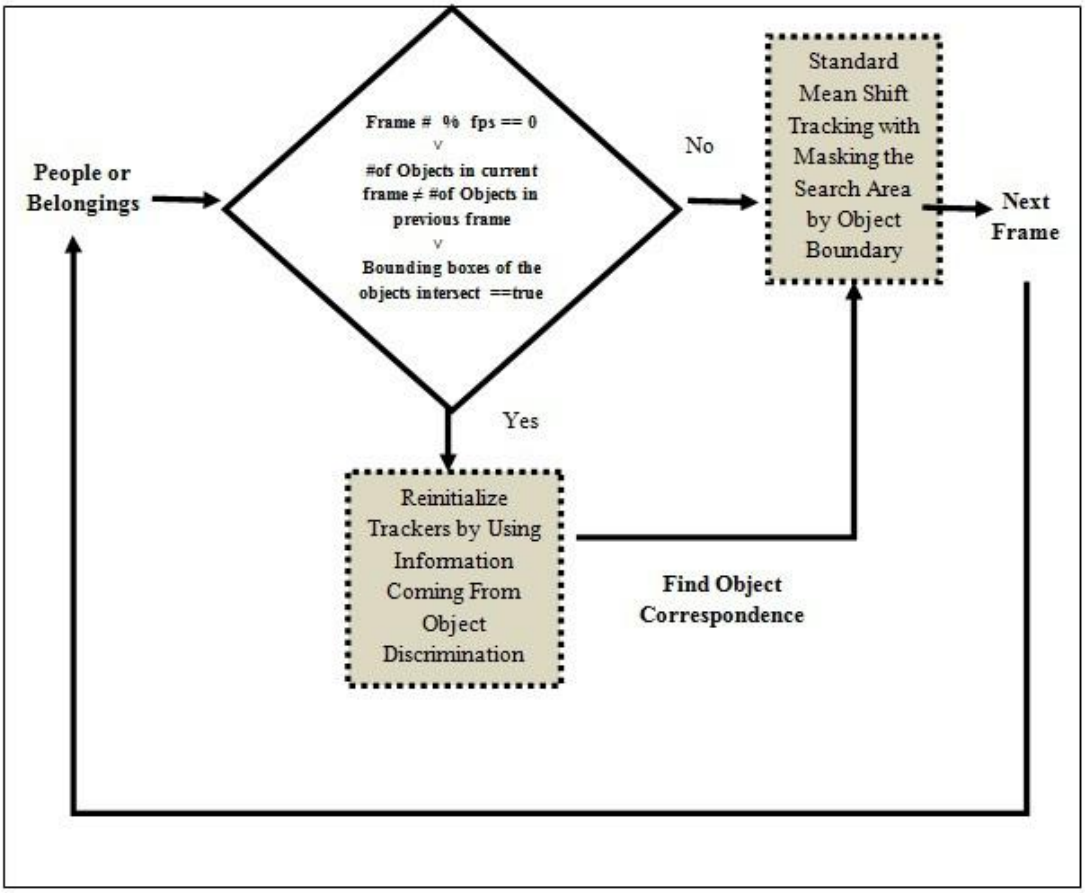


Figure 6

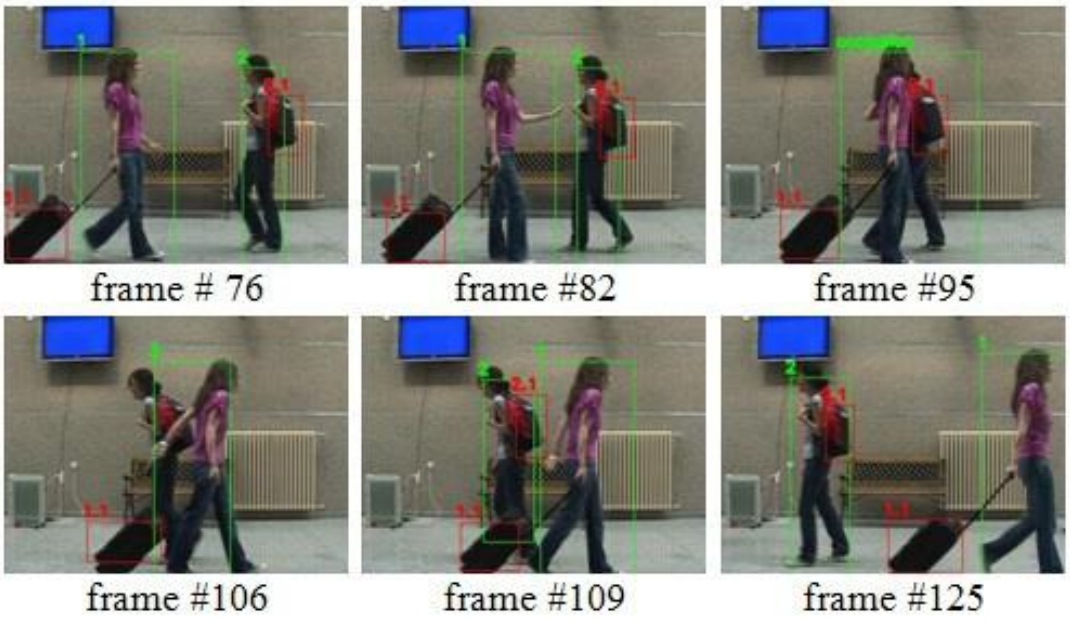


Figure 7

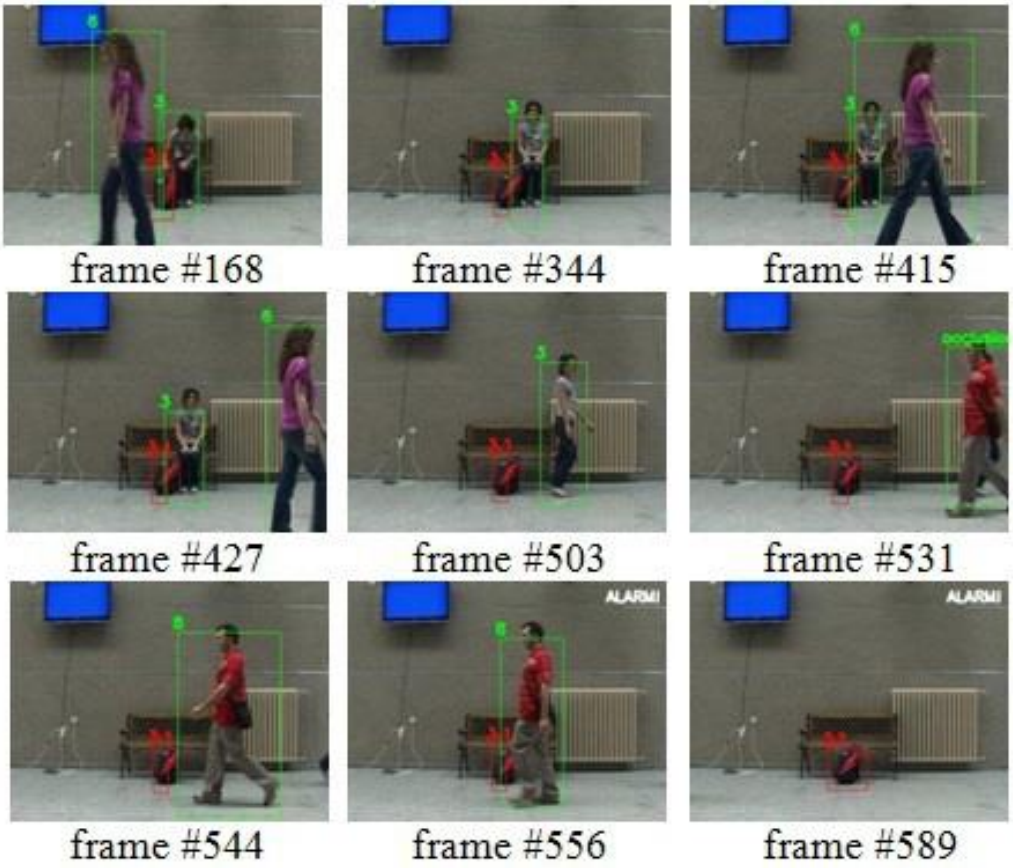

frame #168
frame #344
frame #415
frame #427
frame #503
frame #531
frame #544
ALARM!
frame #556
ALARM!
frame #589

Figure 8

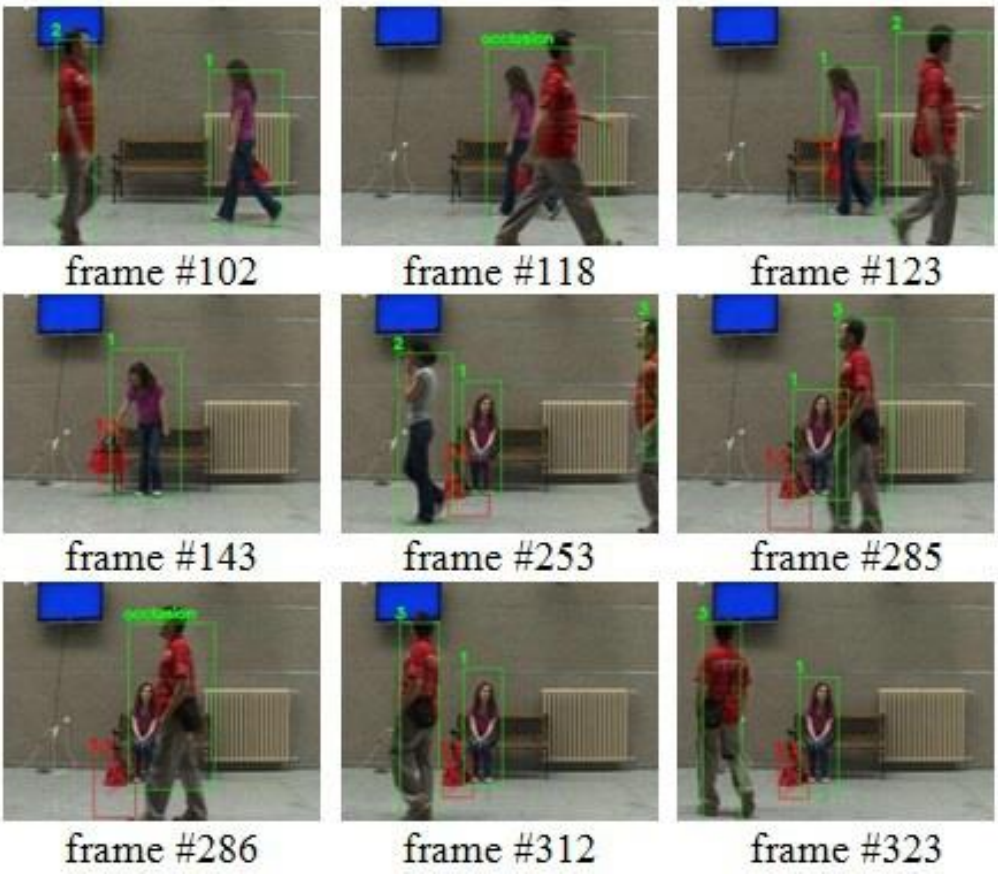
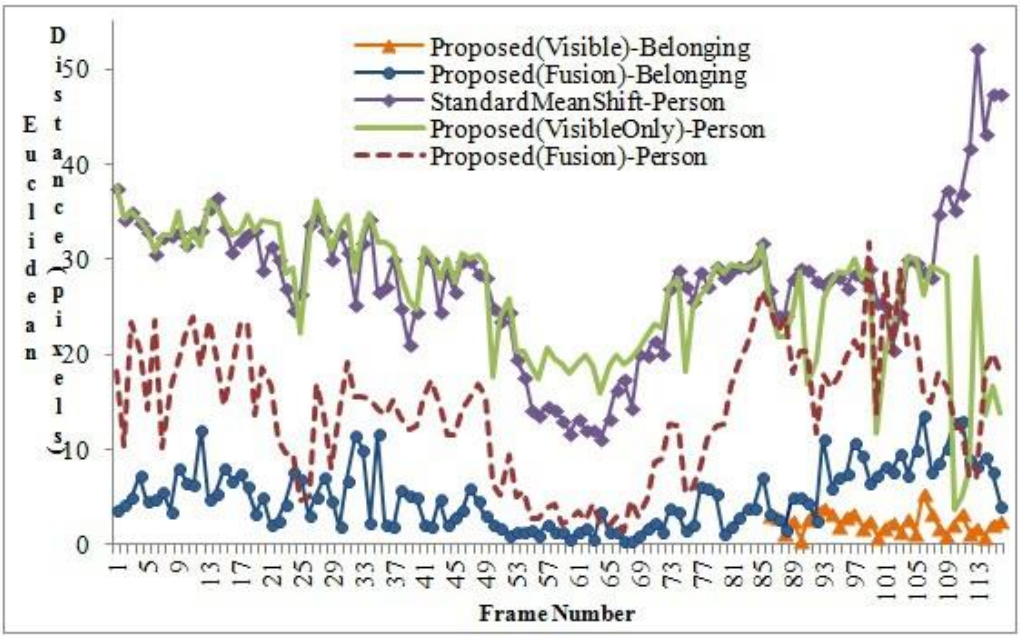


Figure 9



Tables

Table 1

Scenario	No. of Frames	No. of People	No. of Belongings	No. of Alarms	General Description of Scenario
Set1	208	1	1	0	No alarm case, backpack is carried
Set2	191	1	1	0	No alarm case, handbag is carried
Set3	539	1	1	0	No alarm case, luggage is carried
Set4	850	3	1	1	Luggage left unattended, multiple occlusion of people
Set5	590	>5	1	1	Backpack left unattended, multiple occlusion of people and belongings
Set6	1081	4	1	1	Handbag left unattended, multiple occlusion of people and belongings
Set7	341	2	0	0	No alarm case, occlusion exits
Set8	122	1	1	0	No alarm case, backpack is carried, running person
Set9	730	1	1	0	No alarm case, backpack is carried, blocking of belongings
Set10	450	2	2	0	No alarm case, occlusion of people and belongings
Set11	740	3	1	1	Abandoned luggage, Multiple occlusion of belongings after it is left unattended, Heating system on
Set12	1460	>5	>5	0	No alarm case, multiple occlusion of people and belongings, heating system on
Set13	493	1	1	1	Backpack firstly left unattended and then removed
Set14	409	1	1	1	Luggage firstly left unattended and then removed

Table 2

Metrics	Tracking of Person			Tracking of Belonging		
	The Proposed Method	**Visible-band only [29]**	**Standard Mean-Shift Tracking [28]**	**The Proposed Method**	**Visible-band only[29]**	**Standard Mean-Shift Tracking [28]**
Euclidean Error (pixels)	**14.09**	26.56	27.88	**5.02**	*2.41**	NA
Precision	**0.82**	0.59	0.39	**0.79**	*0.95**	NA
Recall	**0.86**	0.83	0.8	**0.93**	0.22	0

Table 3

Scenario	Method [30]		Proposed Method	
	True Alarm Detection	**# of False Alarms**	**True Alarm Detection**	**# of False Alarms**
Set1	NA	**0**	NA	**0**
Set2	NA	**0**	NA	**0**
Set3	NA	**0**	NA	**0**
Set4	**1**	1	**1**	**0**
Set5	**1**	1	**1**	**0**
Set6	**1**	1	**1**	**0**
Set7	NA	**0**	NA	**0**
Set8	NA	**0**	NA	**0**
Set9	NA	**0**	NA	**0**
Set10	NA	**0**	NA	**0**
Set11	**1**	1	**1**	**0**
Set12	NA	**0**	NA	**0**
Set13	**1**	**0**	**1**	**0**
Set14	**1**	**0**	**1**	**0**