Automatic Background Updating for Abandoned Object Detection at Train Stations

Stephen Karungaru, Kenji Terada, and Minoru Fukumi

***Abstract*—In video surveillance using overhead cameras, it is very important to capture and regularly updated the background to enable accurate extraction of persons or objects of interest. However, in scenes with a lot of movement and stationary objects, for example a train station, it is not easy to update the background without including such objects. In this work, we investigate several objects features and combine them to maintain a stable background image. The features include the speed of the object, texture, shape, associations between persons and objects, etc. The data used is a subset of the i-Lids dataset that was captured for analyzing video systems. It is captured in a train station using one overhead camera. Each video segment is about three minutes long.**

***Index Terms*—Active background, abandoned objects, shape matching.**

I. Introduction

Abandoned objects detection is a task to automatically detect and tract objects that are left behind by persons in public areas. Abandoned objects in public areas are a major concern because they might pose a security risk if they contain dangerous materials. Manual detection of such objects is difficult, consumes a lot of manpower and is expensive. To assist the security personnel monitoring live surveillance video, image processing and artificial intelligence methods can be used to automatically detect such objects and alert the monitoring officers to take the appropriate action.

In this area of research, data is not always readily available because of various constrains e.g. privacy, corporate patents, etc. However, some online databases have been offered. On such database for abandoned baggage detection is available on line and has been adapted in this work.

One might ask an interesting question. What is an abandoned object? An abandoned object is defined as follows, using three rules [1],

- Contextual rule: A baggage belongs to the person entering a scene with it.
- Spatial rule: a baggage is unattended if the owner is more than 3 meters away from it.
- Temporal rule: an unattended object that remains in the same area for sixty consecutive seconds is an abandoned baggage.

Manuscript received May 18, 2012; revised August 8, 2012.

The authors are with the graduate School of Advanced Technology and Science, The University of Tokushima, Janpan (e-mail: karunga@is.tokushima-u.ac.jp).

Background subtraction is one of the most commonly used methods for movement detection in videos. However, most areas of interest are sensitive public places and infrastructures that are susceptible to being crowded. Tracking people in a crowded environment is a big challenge, since, in image space, we must deal with merging, splitting, entering, leaving and correspondence [2]. Therefore, a dynamic background, as proposed in this work is vital for such systems.

A. Related Works

As summarized in [1], proposed methods for video event recognition can be divided into three sub-categories based on the underlying features they use: (1) those based on 2-D image features such as optical flow [3] , body shape/contour [4], space-time interest point [5], (2) those based on 2-D trajectories of tracked objects such as hands, legs or the whole body [6] and (3) those based on 3-D body pose acquired from motion capture or 3-D pose tracking system [7]. None of above methods, however, involves the manipulation of objects such as luggage.

II. The Database

The database used in this work was captured for analyzing video systems proposed by multiple researchers [8]. It is captured in a train station using one overhead camera. The camera is attached approximately 40 degrees to the tracks. The movements in the video are as follows,

- Trains continuously are entering and leaving the station
- People moving back and forth, near to, and far away from the camera.
- People getting in and off the trains
- Others waiting for trains or just moving around

The scenes are crowded and very complicated, not only for detecting the people, but also in the detection and maintaining of a stable background.

The datasets for the event detection scenarios each contain approximately 24 hours of footage. Each dataset consists of two or three camera views referred to as stages, and are further segmented into shorter video clips about 3 minutes long.

Fig. 1 shows an example frame extracted from the database.

III. Background Detection

Background subtraction is one of the most commonly used

methods for detecting movement in videos. The advantages of this method include fast processing speeds and ease of detecting new items entering the scene. However, background updating is not a trivial problem, especially for stationary and slow moving objects. In this section we explain in details the algorithm used to detect and update the background.

Fig. 1. An example scene from the database

Initially, the first frame in the sequence is taken to be the background. Thereafter, we use the background subtraction method to detect movements starting with the second frame. Basically, moving objects should not be part of the background. However, this work uses a database captured at a train station with many trains entering, leaving and stopping to drop and pick up passengers.

Therefore, as opposed to other background situations, in this scene, parts of the background are also moving. In this case, not all moving regions detected using the background subtraction, are people. Some moving parts must also be updated as part of the background. Therefore, we must differentiate between the people and the trains.

A. Preprocessing

Every frame must be preprocessed before other procedures are performed. After the background subtraction method is performed the following processes are carried out,

- Image binalization: The threshold is implicitly set at 20.
- A 3x3 median filter is applied to the result to remove noise
- Then, the dilation process is carried out.
- The blobs are extracted using 20 pixels as the neighborhood, Fig. 2.
- Then, in the extracted blobs, further segmentation using contours is carried out. Blobs less than 20x20 pixels in size are ignored. The remaining blobs are the candidate regions, Fig. 3.

B. Object Tracking

When using the background subtraction method, it is not obvious which object to track in the next frame. To solve this problem, one of the common methods used is template matching. That is, capture the present region as the template, and match it to the extracted regions in the following frame to detect its presence. While this method is effective, it is computationally expensive especially when the template image sizes are large.

Therefore, in this work, we propose the use of color histogram as the feature to track. Instead of matching images like in template matching, we match color histograms. Since the size of the histogram bins is fixed and the histograms are normalized, this method is fast and effective in tracking objects. It is also rotational invariant.

Fig. 2. Candidate regions extraction using neighborhood

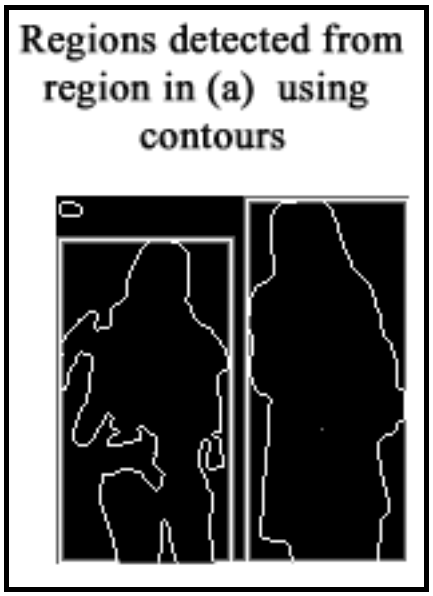

Fig. 3. Candidate regions extraction using contours

The *YCrCb* color space is used to construct the color histograms. This color space is selected because it can also be used to detect skin color regions, which can be further used to detect humans.

The conversion from RGB color space to *YCrCb* color space is performed using the following equations.

$$Y = 0.299*R + 0.587*G + 0.114*B \quad (1)$$

$$Cr = (R-Y)*0.713 + \delta \quad (2)$$

$$Cb = (B-Y)*0.564 + \delta \quad (3)$$

where $\delta = 128$, Y is the brightness, Cr and Cb are the color components.

The tracking process proceeds as follows:

- Capture the first image, and save it as the background
- Beginning the second frame, detect changes in the background using background subtraction
- For each detected region, convert the pixels color to *YCrCb* color space and construct the *Cr* and *Cb* color components histograms
- From the third frame onwards, repeat step 3.
- Match the color histograms in consecutive frames

Fig. 4 shows the calculated histograms. As shown in the figure, the two histograms are very similar even though the images are captured 54 frames apart. This proves that we can use the histograms to track an object in a different scene at high speeds as opposed to template matching which matches the images directly.

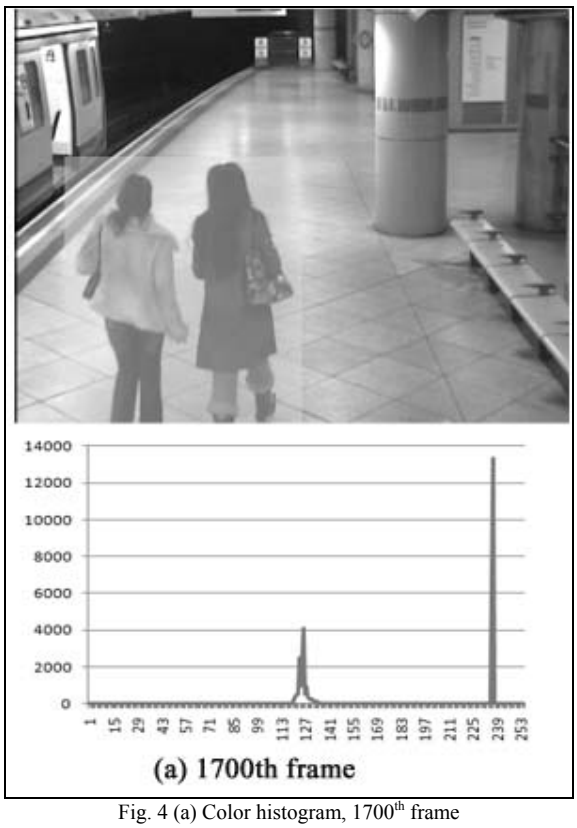

Fig. 4 (a) Color histogram, 1700th frame

A. Background Updating

The color histogram tracking method described above only informs us which objects are moving and in which direction. However, there is no information as to what kind of object they are. In the database used, movement is mostly expected from the trains and humans.

Therefore, we must separate the humans as foreground and trains as part of background. Because of the over-crowding in the scenes and humans walking in and out of the scenes from different directions differentiating between the two is not trivial. We use the following rules to decide which areas are background candidates:

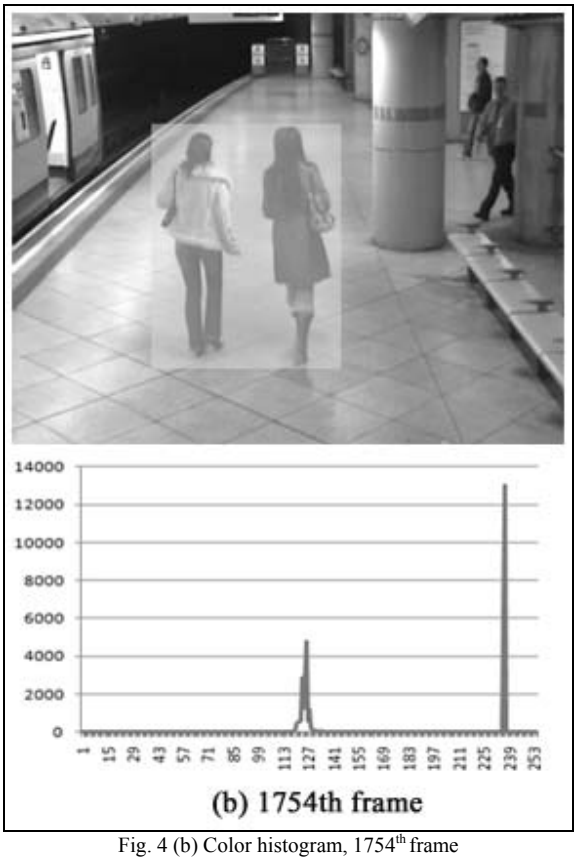

Fig. 4 (b) Color histogram, 1754th frame

- Division: We divide the scene into two regions, near to and far from the camera. Based on this, the trains occupy on the top region (far from camera), while the people can occupy both areas.
- Shape: All regions in the top region satisfying the following equation are part of the background.

$$Width > Height$$

- Direction: Objects in the top region not moving beyond the left half are part of the background.

The areas identified as background must be updated. Moreover, all the 20x20 pixels regions ignored in the preprocessing step are also considered as background and are gradually updated. Note that in this method, not all the background is updated. This greatly improves the processing time.

B. Human Tracking

All other remaining areas, not identified as background, are human candidates regions that must be further sorted to separate the humans from the shadows and luggage they may be carrying.

In this work, since the blobs are extracted using a two-step processing, neighborhood and contours, it is possible to detect multiple humans, something that is normally challenging. The results of these algorithms are as shown in the Fig. 3. Notice that the humans have been individually detected. Fig. 5 and 6 show examples of human detection from the video. The number on the top right is the frame number.

Fig. 5. Human detection

Fig. 6. Human detection

IV. Experiment

To prove the effectiveness of this work, we carried out experiments. The data used in this work is a three minute video consisting of about 5400 frames.

The first frame is taken as the background and is updated using the methods described in this work. The accuracy of this work is calculated based on the number of frames in which the human subjects were accurately detected and on the processing time. Three experiments were carried out by varying the number of frames per set as follows:

- Set 1: 1 frame (30ms)
- Set 2: 6 frames (180ms)
- Set 3: 12 frames (360ms)

The difference in the experiments is in the number of frames processed. For example, in set 1, we process all frames while in set 2, we process every other sixth frame.

V. Results

The results of the three experiments are shown in Table I.

As shown in Table I, the results in both the accuracy and the processing time improved as the number of images processed reduced.

TABLE I: Processing Time Using the Proposed Tracking Method

Experiment	Accuracy (%)	Time per set (average in seconds)
Set1	93	0.027
Set2	95	0.020
Set3	98	0.017

VI. Discussion

We think the reason to be that, as the gap between processed frames increases, it is easier to detect the direction of movement since there are less overlapping. However, the non-human region increases and requires more time to update the background as compared to smaller gap being processed.

VII. Conclusion

In this paper, we proposed an automatic background updating method that is fast and robust for video processing applications especially surveillance systems. Background subtraction method, coupled with color histogram matching method was used to detect the active regions and track them respectively. However, because not all active regions could be considered part of the background, shape analysis and movement direction were used to distinguish between human and train areas.

To prove the effectiveness of the proposed method, experiment using a three minute video (about 5400 frames) captured at a train station was used to test the method. The results show that the background was actively and accurately updated and that all human areas were also accurately detected.

In the future, we will extend this work to abandoned objects detection by further analyzing the human regions detected to establish when they leave objects in their possession for more than sixty consecutive frames.

References

[1] Fengjun Lv, Xuefeng Song, Bo Wu, Vivek Kumar Singh,and Ramakant Nevatia, Left-Luggage Detection using Bayesian Inference, 9th Intl. Workshop on Performance Evaluation of Tracking and Surveillance (PETS-CVPR'06), June 2006

[2] Dahmane and Jean Meunier, "Left-luggage detection using homographies and simple heuristic"s, IEEE international workshop on performance evaluation in tracking and surveillance(PETS), 2006.

[3] A. A. Efros, A. C. Berg, G. Mori and J. Malik. "Recognizing Action at a Distance". In IEEE International Conference on Computer Vision 2003, pp.726-733.

[4] A. Yilmaz and M.Shah. "Actions Sketch: A Novel Action Representation", In IEEE Conference on Computer Vision and Pattern Recognition 2005, pp.984-989.

[5] I. Laptev and T. Lindeberg. Space-time interest points, In IEEE International Conference on Computer Vision 2003, pp.432-439.

[6] C. Rao, A. Yilmaz and M. Shah. View-Invariant Representation and Recognition of Actions, In International Journal of Computer Vision 50(2), Nov. 2002, pp.203-226.

[7] F. Lv and R. Nevatia. "Recognition and Segmentation of 3-D Human Action using HMM and Multi-Class AdaBoost" In European Conference on Computer Vision 2006, pp. 359-372

[8] i-Lids dataset for 2007 IEEE International Conference on Advanced Video and Signal based Surveillance, http://www.eecs.qmul.ac.uk/andrea/avss2007d.html, 2011.

Stephen Karungaru received a PhD in Information System Design from the Department of information science and Intelligent Systems, University of Tokushima in 2004. He is currently an associate professor in the same department. His research interests include pattern recognition, neural networks, evolutionary computation, image processing and robotics. He is a member of IACSIT, RISP, IEEE and IEEJ.

Kenji Terada received the doctor degree from Keio University in 1995. In 2009, he became a Professor in the Department of Information Science and Intelligent Systems, University of Tokushima department. His research interests are in computer vision and image processing. He is a member of the IEEE, IEEJ, and IEICE.

Fukumi received the doctor degree from Kyoto University in 1996. Since 1987, he has been with the Department of Information Science and Intelligent Systems, University of Tokushima. In 2005, he became a Professor in the same department. He received the best paper award from the SICE in 1995. His research interests include neural networks, evolutionary algorithms, image processing and human sensing. He is a member of the IEEE, SICE, IEEJ, IPSJ, RISP and IEICE.