IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 3, June-July, 2014
ISSN: 2320 – 8791 (Impact Factor: 1.479)
www.ijreat.org

Automated People Tracker For Visual Surveillance Application

Vishal Ganvir[1], Manohar Buktare[2], Dr. S. L. Nalbalwar[3]

[1,2,3] **Department of Electronics and Telecommunication, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, Maharashtra.**

Abstract

Moving object detection and tracking is often the first step in applications such as video surveillance. The main aim of project is moving object/people detection and tracking system with a static camera to provide a system that tracks particular person in large number of video clips and gives us a single video clip consisting of several video clips combined together. We propose a general moving object detection and tracking based on vision system using image background subtraction algorithm. This paper focuses on detection of moving objects in a scene for example moving people talking with each other, and tracking and detection of people as long as they stay in the scene. This is done by background subtraction algorithm with the help of Simulink in MATLAB software.

In this paper we estimated the position of moving people and tag them by particular Id. And then this Id is used to identify them in other videos captured by multiple camera networks.

Keywords— *computer vision system, moving object detection and tracking, background subtraction algorithm, Kalman filter.*

1. INTRODUCTION

Video surveillance of human activity usually requires people to be tracked. It is important to security purpose and traffic control which is also used to take necessary step for avoiding undesired interaction [1].

We present our system for moving object detection (People) and tracking using a static webcam that monitors a typical open work area. Object tracking is central to any task related to vision systems. We present a vision system for moving people detection and tracking therefore taking video at no change of illumination area with particular background and in this background people are moving.

In the process leading from an acquired image to the information about objects in it, two steps are particularly important: foreground masking and tracking [2]. In this paper we present a simplified detection method based on background subtraction and a blob matching tracking algorithm that relies only on blob matching information and KALMAN filter which uses statistical descriptions to model and predict motion characteristics using MATLAB software.

A moving object detection and tracking system with a static camera has been developed to estimate position parameters. For that application we presented background subtraction algorithm which contain subtraction between reference image (background image) and number of images captured by a static camera and we also need some morphological operation, calculation of centroid of blobs and distance between moving object in scene.

The paper is organized as follows: An overview of the related research is given in Section 2. The proposed system overview is presented in Section 3. Background subtraction algorithm is explained in Section 4. Section 5 and 4 gives the simulated results and its analysis. Finally the project is concluded in Section 7.

2. RELATED WORK

Real time automated visual surveillance has been a popular area for scientific and industrial research since it was pioneered by O'Rourke and Badler (1980) and Hogg (1983). People tracking naturally plays a key role in any visual surveillance system, and a number of tracking algorithms for different applications have been presented (Baumberg, 1995; Br´emond and Thonnat, 1997; Cai et al., 1995; Gavrila and Davis, 1996; Haritaoglu et al., 2000; Johnson, 1998; Khan et al., 2001; Lipton et al., 1998; Sidenbladh et al., 2000; Wren et al., 1995); Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching' (2011).

Using the classification scheme defined in Section above, the tracking algorithms can again be classified into the three main categories of increasing model complexity:

Category 1: Methods using region- or blob-based tracking: Br´emond and Thonnat (1997); Cai et al. (1995); Khan et al. (2001); Lipton et al. (1998); Wren et al. (1995).

Category 2: Methods using a 2D appearance model of a human being: Baumberg (1995); Haritaoglu et al. (2000); Johnson (1998).

Category 3: Methods using an articulated 3D model of a human being: Gavrila and Davis (1996); Sidenbladh et al. (2000).

In the present work the concepts of continuous template matching and frame differencing have been used to implement a robust automated moving object/people tracking system. In this implementation a web camera of laptop has been used to grab the video frames and track an object. Using frame differencing on each consecutive frame a moving object is detected with high accuracy and efficiency. Once the object has been detected it is tracked by employing an efficient Template Matching algorithm and Kalman filter. The templates used for the matching purposes are generated at that time only. This ensures that any change in the pose of the object does not hinder the tracking procedure. To automate the tracking process the camera is mounted on a pan-tilt arrangement, which is synchronized with a tracking algorithm. When the object is being tracked moves out of the viewing range of the camera, then pan-tilt setup is automatically adjusted to move the camera, so as to keep the object in view. It works on following algorithm that includes some steps for moving camera by pan tilt system.

1. Take current image and previous image.
2. Take difference between them.
3. Select thresholding.
4. Difference of image is greater than threshold object is detected.
5. Find centroid of detected object.
6. Generate template and take coordinate of template.
7. (Template matching algorithm)

    IF the template matching is successful

      THEN

        IF the tracker has NOT detected motion of the object

          AND the detector has

          THEN goto STEP 1 (get a new template)

        ELSE goto STEP 5 (get the x, y position)

    ELSE goto STEP 1 (get a new template)

8. Obtain the position P(x, y) of the match and pass it on to the pan-tilt automation module for analysis.
9. Get the direction of horizontal and vertical movement of tracked object.
10. On based of movement within certain dimension it decide movement of camera in clockwise or anticlockwise
11. Else go to step 1.

The more detailed the models for people detection and tracking, the better the system can handle the particular situations for which it is trained. However, systems in Category 3 with complex 3D models, e.g. Sidenbladh's 3D People Tracker, are currently too slow to use in real-time systems. They might also require a special camera/scene setup. It requires a very robust and precise motion detector. Given that, it works very well and correctly tracks and labels people. These 2D methods are fast enough for use in a real-time system.

3. SYSTEM OVERVIEW

The presented system contains vision system that can capture videos and other is background subtraction algorithm that can process for moving object detection and tracking.

For many vision-based systems, it is important to detect a moving object automatically. Image processing, analysis, and machine vision represent an exciting and dynamic part of cognitive and computer science. For this project, the Vision System which is used includes a webcam of a laptop.

Vision system pc should be having frame grabber card which is support to camera and it should has fast processor for capturing frame with snapshot.

The system overview is shown in figure 1. The video is captured from image acquisition system. Reading of all images or frame is done in Matlab. The first image is called background or reference image. All the other following images are subtracted from background, then if difference is greater than threshold the object is detected.

For tracking side we used some simulink tollbox of matlab with properties of centroid, bounding box and area of white pixels. Using the bounding boxes we show tracking of moving object.

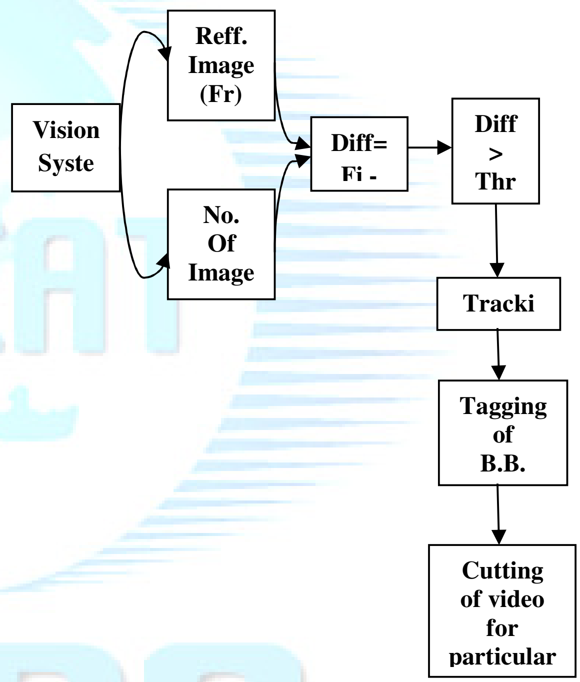


Figure 1 System Overview

4. BACKGROUND SUBTRACTION ALGORITHM

There are various techniques for moving object detection and tracking like optical flow, low change of illumination, segmentation, background subtraction, frame difference etc. We formulated the problem in a sequential manner. We planned different steps with different set of operation that will be taking place at each step and the output of that step will be used as the input to the other step. Each step is in charge of specific function which it will be performing on each frame of the video sequence and the final result of that step will be used in another step and each step will follow the same things. The last step will give the final output in the form of a video in a well structured way. The formulation of step are defined as follows-

1. Take video from Vision System.

2. Read 1st image that is reference image (Background).
3. Read the following next images.
4. Take subtraction of them and compare with threshold value.
5. Apply morphological operation like dilation and erosion for small noise removal.
7. Analyse the blobs in the video which are related to same target and merge them together.
9. Then draw rectangle plot according to the centroid of that merged blobs that is bounding box for that target
10. Assign particular Id to that bounding box.
11. Then using these measurements and Kalman filter track that bounding box in the video.
12. The system will note down the time of entering the particular bounding box and leaving too.
13. When particular Id is given to system it trims the whole video to the time for which that bounding box with the specified Id is present in the video.

The actual system developed in Simulink using vision system toolbox, image acquisition toolbox, DSP system toolbox, math operation, signal routing toolbox, etc.

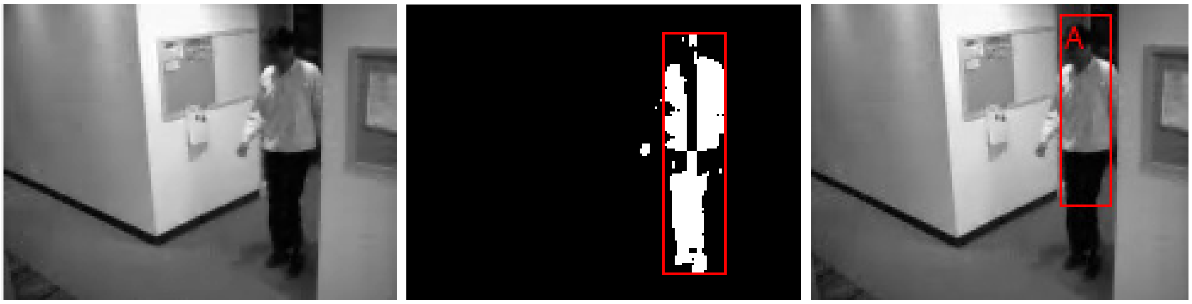
5. SIMULATED BACKGROUND SUBTRACTION ALGORITHM

We have simulated background subtraction algorithm in MATLAB with a video already stored in workspace. The video is about man moving in particular area. He will be detected and tracked using the simulation of algorithm. The results are shown below with different time steps. We have also simulated this algorithm for real-time videos and the results are quiet good with 85 percent efficiency.

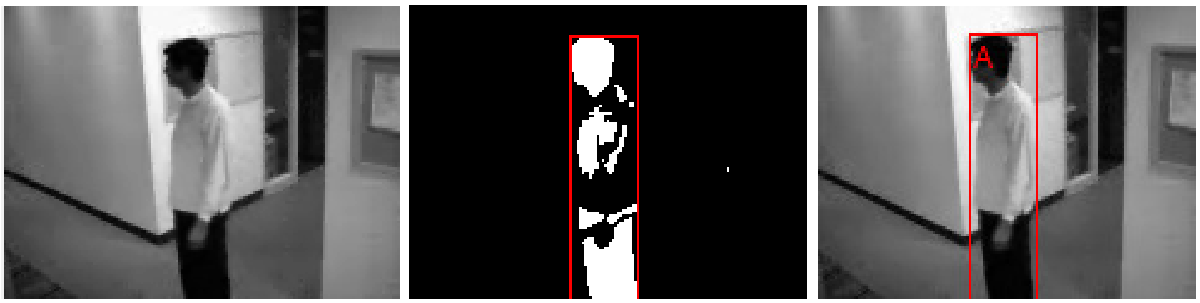
Figure 2 shows the first image of video at time $t = t_0$ which is treated as reference image or background image.

Figure 2 Reference Image (Background Image)

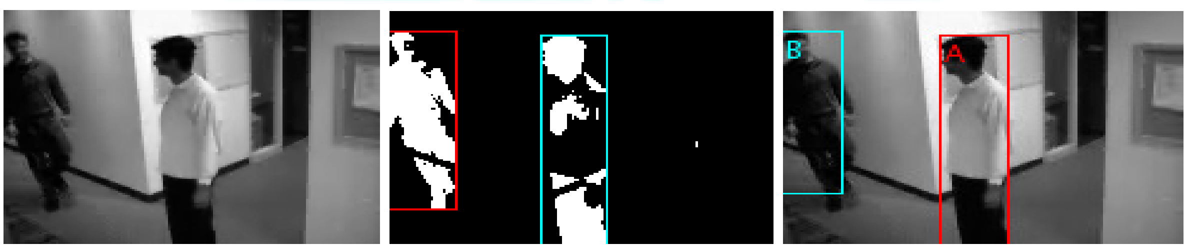
All these results are for static background.
Figure 3 shows the number of images following the reference image in video. The first image is actual image at time $t = t_1$, next is binary image output of subtraction of image at time t_1 and background image at time t_0 and the pixels which are different are treated as blobs and these blobs are indicated by building bounding box around it. The next is actual image in



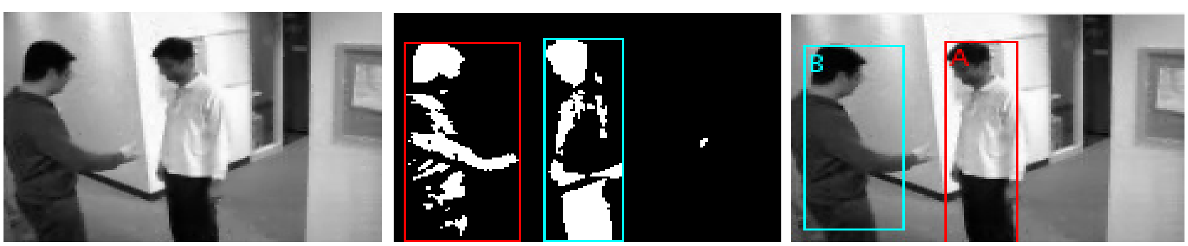
a) At time t= 1.7 sec



b) At time t= 2.7 sec



c) At time t= 3.4 sec



d) At time t= 4.5 sec

Figure 3 Simulated results showing detection and tracking of people

video stream with bounding box surrounding the moving object or person. In the figure below there are some results of detection and tracking of moving object or person at different time stamp.

The figure below shows the graph of bounding box co-ordinates with respect to time. From this we can assign time interval which we need to cut from the original video. The graph of both detection system and tracking system is shown in figure 4a and 4b. From these graphs we can clearly see that the entering time of first person is 0.8 sec and for second person it is 2.9 sec after the video is started. These time stamps are used to cut the original video to get a clip of specific person.

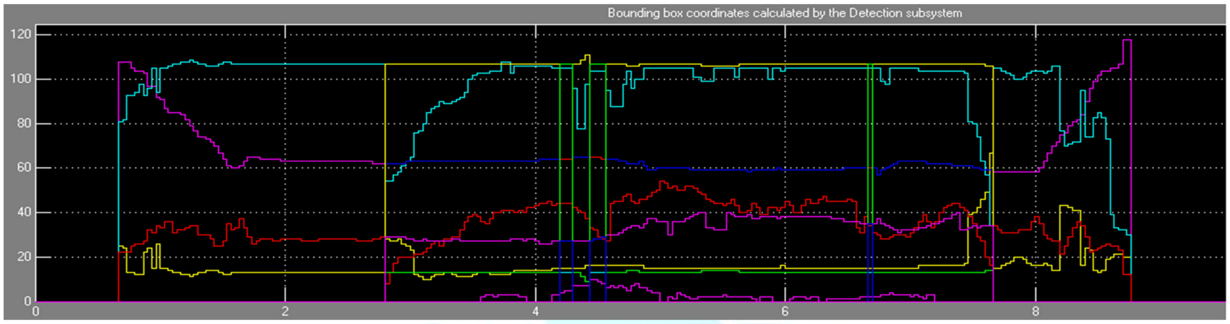


Figure 4a Detection system results

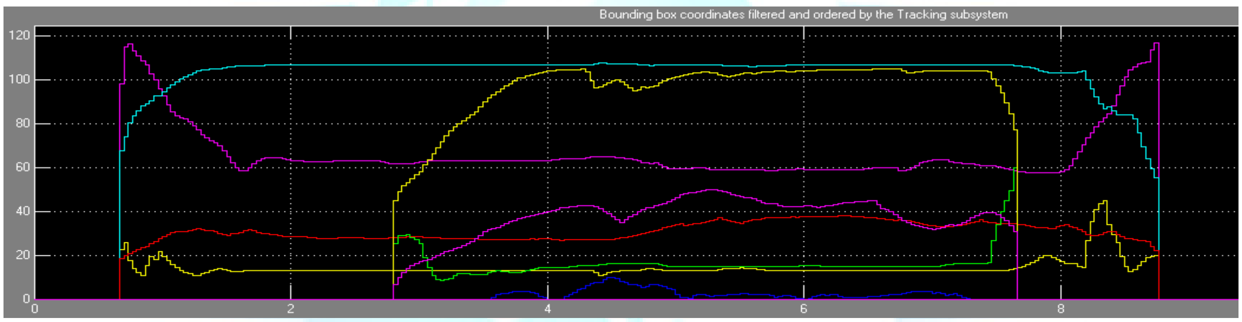


Figure 4b Tracking system results

6. RESULT ANALYSIS

The results are calculated and analysed and it is found to be 85 percent accurate. The binary image of foreground is given to the blob analysis block of computer vision system toolbox. It gives us the total number of blobs present in the image and their co-ordinates.

At time t= 3.4 sec the output of blob analysis is, total number of blobs i.e. num = 5 and their co-ordinates are

$$\text{Pos} = \begin{bmatrix} 11 & 13 & 83 & 32 & 81 & 0 & 0 & 0 \\ 0 & 63 & 67 & 81 & 81 & 0 & 0 & 0 \\ 94 & 49 & 37 & 10 & 8 & 0 & 0 & 0 \\ 36 & 19 & 23 & 9 & 10 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{And Enable} = [1\ 1\ 1\ 1\ 1\ 0\ 0\ 0]$$

Enable is active for only those columns in which position co-ordinates are present otherwise it is set to zero.

But actually there are only two persons so it should show only their co-ordinates. Then next step is to merge the blobs which belong to same target. This is done by the subsystem "merge blobs belonging to the same target". The pos co-ordinates are compared column-wise and then blobs which are intersecting are merged together and new co-ordinates are formed and other columns are replaced by zero column.

The output of detection block is new set of position co-ordinates and respective enable output.

$$\text{Pos\_detect} = \begin{bmatrix} 11 & 13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 63 & 0 & 0 & 0 & 0 & 0 & 0 \\ 94 & 107 & 0 & 0 & 0 & 0 & 0 & 0 \\ 36 & 28 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{En\_detect} = [1\ 1\ 0\ 0\ 0\ 0\ 0\ 0]$$

From this we conclude that the position of first blob doesn't intersect with any other blob and 2nd, 3rd ,4th and 5th blobs intersect somewhere and hence their new position co-ordinates are calculated. All other columns are replaced by zero. All these calculations are carried on the binary image so it may have some noise or quantization error. This error or noise is removed with the help of Kalman filter.

We use the Kalman Filter block to predict or estimate the state of a dynamic system from a series of incomplete and/or noisy measurements. Suppose you have a noisy linear system that is defined by the following equations:

$$x_k = Ax_{k-1} + w_{k-1}$$

$$z_k = Hx_k + v_k$$

This block can use the previously estimated state, $\widehat{x_{k-1}}$, to predict the current state at time k, $\overline{x_k}$, as shown by the following equation:

$$\overline{x_k} = A\widehat{x_{k-1}}$$

$$\overline{P_k} = A\widehat{P_{k-1}}A^T + Q$$

The block can also use the current measurement, z_k , and the predicted state, $\overline{x_k}$, to estimate the current state value at time k, $\widehat{x_k}$, so that it is a more accurate approximation:

$$K_k = \overline{P_k}H^T(H\overline{P_k}H^T + R)^{-1}$$

$$\widehat{x_k} = \overline{x_k} + K_k(z_k - H\overline{x_k})$$

$$\widehat{P_k} = (I - K_k H)\overline{P_k}$$

The variables in the above equations are defined in the following table.

Variable	Definition	Initial value or condition
X	State	N/A
$\hat{x}$	Estimated state	`zeros([6, 1])`
$\bar{x}$	Predicted state	N/A
A	State transition matrix	$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
W	Process noise	N/A
Z	Measurement	N/A
H	Measurement matrix	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
V	Measurement noise	N/A
$\hat{P}$	Estimated error co-variance	10*I(6)
$\bar{P}$	Predicted error co-variance	N/A
Q	Process noise co-variance	0.05*I(6)
K	Kalman gain	N/A
R	Measurement noise co-variance	2*I(4)
I	Identity matrix	N/A

Table 1 Variables used in equations

The output of Kalman filter is,

$$\text{Pos\_track} = \begin{bmatrix} 11.09 & 13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 63 & 0 & 0 & 0 & 0 & 0 & 0 \\ 85.09 & 107 & 0 & 0 & 0 & 0 & 0 & 0 \\ 27.35 & 28.07 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{En\_track} = [1 \;\; 1 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0]$$

Next is to plot bounding box in video with the position co-ordinate and enable signal.

From the graph of position against time we get time co-ordinates of any bounding box i.e. their entering time and leaving time and using this time stamps we can cut the original video to get video clip of any particular person.

Person	Time of entrance	Time of leaving
A	0.8 sec	8.7 sec
B	2.9 sec	7.9 sec

Table 2 Entrance and Leaving times

After giving this co-ordinate to cutter and selecting person A, it will give (8.7 - 0.8) 7.9 sec clip starting from 0.8 sec of original video.

7. CONCLUSIONS

We have presented and implemented automated people tracker with the help of background image subtraction algorithm in Matlab. After simulation, we got nice result compare to other research by using Kalman filter to remove noise, and quantization error, also by using morphological operation such as dilation.

We have done detection, tracking and estimating time of entrance and leaving of any particular person. We also showed these results graphically. We are getting nearly 85 percent efficiency of this system. The cutting of video for tracking any particular person is also done.

ACKNOWLEDGMENT

This is a great opportunity for me to thank my project guide, Dr. S. L. Nalbalwar, Head of the Department of Electronics and Communication Engineering for their valuable guidance and kind support for the project. I am very fortunate that I got an opportunity to work under him and to learn from him. He has always inspired me to give my best in project work. I, also thank all the teachers and staff members of Department of EXTC for their kind support. I express my grateful thanks to my classmates and colleagues for their kind help, encouragement and support whenever needed.

References

[1] Q. Cai and J. K. Aggarwal. "Tracking human motion in structured environments using a distributed-camera system". *IEEE Tansactions On PAMI*, 2(11), Nov 1999.

[2] T.Chang and S. Gong. "Tracking multiple people with a multicamera system". In *IEEEWorkshop onMulti-Object Tracking*, 2001.

[3] Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching'Karan Gupta1, Anjali V. Kulkarni2 Indian Institute of Technology, Kanpur, India

[4] C. Smith, C. Richards, S. Brandt, N. P. Papanikolopoulos, "Visual Tracking for Intelligent Vehicle Highway Systems," *IEEE*

Transactions on Vehicular Technology, Vol. 45, No. 4, Nov. 1996, pp. 744-759.

[5] S. Dockstader, A. Tekalp, "Multiple Camera Fusion for Multi-Object Tracking," *Proc. of IEEE Workshop on Multi-Object Tracking*, pp. 95-102, 2001, Vancouver, BC, Canada.

[6] A. Tyagi, G. Potamianos, J. Davis, S. Chu, "Fusion of Multiple Camera Views for Kernel-based 3D Tracking," *Proc. of IEEE Workshop on Motion and Video Computing*, Feb. 2007, pp. 1-6, Austin TX.

[7] Adarsh Kowdle and Tsuhan Chen, "Learning to Segment a Video to Clips Based on Scene and Camera Motion",Cornell University, Ithaca, NY, USA,2012.