Parallel Kalman Filter-Based Multi-Human Tracking in Surveillance Video

Abdul-Lateef Yussiff, Suet-Peng Yong, Baharum B. Baharudin
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, Tronoh
Perak, Malaysia.
MALAYSIA
ayussiff@gmail.com, yongsuetpeng@petronas.com.my, baharbh@petronas.com.my

***Abstract*—A novel approach to robust and flexible person tracking using an algorithm that integrates state of the arts techniques; an Enhanced Person Detector (EPD) and Kalman filtering algorithm. This proposed algorithm employs multiple instances of Kalman Filter with complex assignment constraints using Graphics Processing Unit (GPU-NVDIA CUDA) as a parallel computing environment for tracking multiple persons even in the presence of occlusion. A Kalman filter is a recursive algorithm which predict the state variables and further uses the observed data to correct the predicted value. Data association in different frames are solved using Hungarian technique to link data in previous frame to the current frame. The benefit of this research is an adoption of standard Kalman Filter for multiple target tracking of humans in real time. This can further be used in all applications where human tracking is needed. The parallel implementation has increased the frame processing speed by 20-30 percent over the CPU implementation.**

***Index Terms*—Human Tracking; Kalman Filter; Multi-person Tracking.**

I. Introduction

The need to increase the installation of surveillance cameras in public places has increased significantly especially since the event of terrorist attack in New York, London, Madrid and other places. Despite, Surveillance system being one of the most active research areas in the computer vision, there has not been much improvement in automating the surveillances due to complexity of the surveillance problem. The increase in computer power recently has further elevated the interest in this field. However, video surveillance's effectiveness and response to an event is mainly determined not by the technological capabilities of the device but by the alertness of the operator monitoring the camera system [1]. The key objectives common to many of the surveillance systems as reported in the literatures are to detect, classify and track object of interest within the surveillance settings.

The problem of human tracking can simply be stated as given sequence of images, can one estimate the positions and other relevant information of person in the sequence. According to [2] tracking problem is deceptively simple and easy to formulate, however the real world implementation is hard due to reasons such as non static and complex background model, foreground object similarity to the background, illumination problem, camera motion, and so on. This problem has been solved by different techniques, among them is the popular feature-based methods or similarity filtering [3]. The feature- based technique first preprocessed the scene to extract relevant representative features of the potential target object, then, feature analysis to make a decision on the target. The drawback of feature-based technique is that performance relies heavily on the ad-hoc decision which are often not optimized [3]. A real time tracking method must be capable of robust performance, should aim at model parameter minimization and its computational complexity should be minimum as much as possible and also should be able to track multiple objects at the same time. Simultaneously tracking multiple persons in video surveillance can be solved by using an estimation method for predicting the position of the human in the subsequent frame using probabilistic Bayesian technique called Kalman Filter on the NVDIA General Purpose Graphical processing Unit (GPGPU). The contribution of this paper to the research communities is two folds. Firstly, real-time tracking of multiple human in video surveillance using Kalman Filter on parallel architectures, Secondly, using a 6 dimensional input feature vector to the Kalman Filter as the measurement Parameters helps improves the tracking accuracy.

Among the application areas are Human Computer Interaction (HCI), Surveillance, Robotics, Ambient Intelligence, Control and Driver-less self driving vehicle. The ability of being able to detect and track human and their activities in a video sequence is of utmost importance for every surveillance system.

This research paper uses multiple instances of Kalman Filter algorithm, implemented on parallel architecture GPU, to predict the states of target human in a video frame. Section 2 briefly addresses the previous works, System overview is discussed in section 3, section 4 addresses people detection. Section 5 and 6 discuss Kalman filter and Tracking Modeling. Experiment and results is discuss in section 7 and finally conclusion.

978-1-4799-0059-6/13/$31.00

II. Previous Works

This problem of simultaneously tracking multiple human has been a subject of research for quite some time now. Significant amount of work has already been reported in the field with each system employing its own unique tracking technique. Among the proposed techniques are: Background subtraction, Mean-shift, Optical flow, Feature matching, Particle Filters, etc [4]. Each proposed algorithm has its weakness as well as its strength.

Background subtraction is used for moving object detection whenever there is static or stationary background which can be used as a reference model. Several earlier works on people detection and tracking technique rely on the background subtraction. Background subtraction techniques generally determined foreground object from the video and then group it into categories like human and non-human depending on color (skin color), contour, shape, or motion and others in a pixel-wise manner. High sensitivity to background changes and illumination, and unsuitability to high density of persons, static camera assumptions, reference model requirement are among the drawbacks of this approach. Optical flow determines quantity of image movement within a time frame. It is used to segment a moving object from its background under the condition of significant difference in velocity between the moving object and the background model.

Haritaoglu et al. [5] employs an area overlap as criterion to find association between objects in different frames in outdoor environment by combining stereo computation into an intensity based detection and tracking. Behrend et al. [6] proposed a technique of tracking uncooperative moving objects using Probabilistic Multiple Hypothesis Tracking (PMHT) on a relational database for storing the tracked data to achieve a real time performance. The authors claimed an efficient performance over the classical PMHT. The problem associated with this kind of technique is inheritance of all problem associated with relational database which can be very slow hence, reducing the speed and this makes real time implementation unachievable. [7] used particle filter for multi-target tracking on confidence Maps by including target ID in the particle state to deal with large and unknown number of targets a priori. ID are assigned to targets by using a Mean shift clustering supported by Gaussian Mixture Model. Particle Filter for tracking [8]–[10] became popular of late because of its robustness in tackling the non linear and non Gaussian problems however it suffers from sample impoverishment problems [11] and high computational cost makes the Particle filter unsuitable for real time tracking. Because of that [11] proposed a Particle Swarm Optimization technique to tackle the problem of sample impoverishment.

III. Proposed System Overview

The architecture is made up of several modules as shown in Fig. 1 and this include: Data preparation, Human Detection, and Tracking. First step is to process and convert the video and images to the format that is compatible with the Matlab and toolboxes. Then the detection of human and extraction of key features from the video frames. Having detected and extracted the relevant features, Surveillance system usually monitor and track object of interest in the environment. Aim of tracking is to monitor the interested object over time by identifying its position in the Cartesian coordinates.

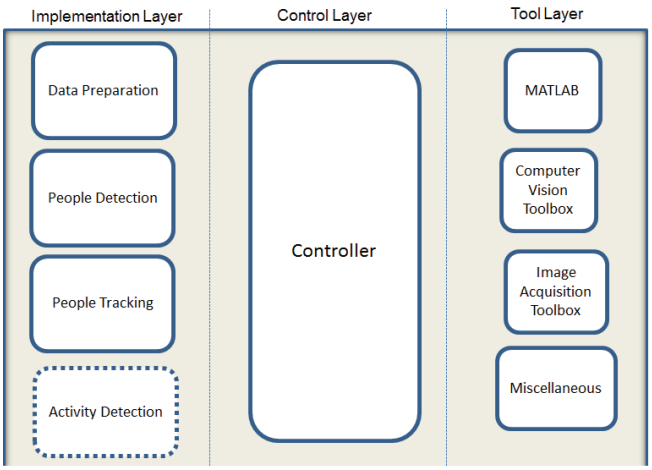

Fig. 1. System Architecture of the framework

A. Data Preparation and Image Source

Most available dataset are compressed using different codecs and hence are not fit to be used. We need a mechanism of converting the downloaded video to the format that will be usable in the Matlab environment. The videos were converted to raw videos and saved in Audio Video Interleaved (.avi) files using ffmpeg [12]; an open source software. Square root gamma correction is applied to the videos to improve its brightness and also to make sure all videos have similar standard characteristics.

B. People Detection

Human detection identifies area of interests for tracking. The key to successful tracking is robust feature extraction and detection. This detection process applies machine learning technique to compare previously trained model on the features to the images, then classify the detected object in terms of human/non- human based on their similarities to the trained model. The detection framework has been discussed extensively in [13]. The adopted technique defines people detection as an hybrid of HOG-based [17] full upright person and the Haar-based upper body [14]. The algorithm detects full upright person in video sequence as well as detecting partial occluded person from the head to chest. For every person detected, the bounding boxes of the person with the following attributes: position in (x, y) in the 2-D Cartesian coordinate, the length, l and the width, w of the rectangular box that contains the human are extracted.

C. People Tracking

Having detected human in the surveillance video, the tracking module tracks the person from frame to frame to identify the position of the person of interest in each frame. The algorithm associates human positions in consecutive frames. Track modeling and the algorithm will be discussed further in this paper.

D. Human activity detection

Aim of every surveillance is not just to detect and track the human but also to detect and recognize their activities. I shall be investigating how to detect activities of a tracked human in the video surveillance in the near future.

IV. Tracking with Kalman Filter

Problem of tracking is to instantaneously determine a prediction about the state (location, velocity, shape, and other relevant information) of an object from a sequence of images. Several different algorithms have been proposed for people tracking. it is evident that there is no single best model for tracking and the success depend not only on the type of technique being used, but also on the nature of data being handled. A detection-to-track technique is employed in this research.

In surveillances, the data are sequentially arriving in time and this is generally a common norm to many real time tracking system. There is a need to predict, and update the prediction as new data are observed, based on the output of the detection module of each frame. Thus, the goal is to find a method for estimating the state vector of the moving person at time t, and also for updating the estimate as new observation becomes available in the next immediate future with minimum prediction (estimation) error and be practically feasible to the problem at hand. Kalman Filter provides the mechanism to achieve this goal. Kalman filter is a mathematical tool used to estimate linear quadratic problems; the problem of estimating the instantaneous state of a linear dynamic system corrupted by white Gaussian noise, using measurements linearly associated to the state [15]. It is expressed mathematically using equations 1 and 2.

$$x_t = Ax_{t-1} + Bu_t + w_{t-1} \tag{1}$$

$$z_t = Hx_t + v_t \tag{2}$$

Equation 1 is called process or state equation while equation 2 is the measurement equation. x_t, system state vector, is a linear combination of its immediate previous state value and a control input u_t together with Gaussian noise w_t; which is called process noise. t is the discrete time step or frame sequence index, Also measured value z_t is a linear combination of state equations and random Gaussian noise called measurement noise. All the noises are statistically independent of each other. $P(w_t) \sim N(0, Q)$ and $P(v_t) \sim N(0, R)$

The variables A, B and H in equations 1 and 2 are model matrices and assumed to be constant. Also the noises w_t and v_t are independent and identically distributed with zero means and covariance.

In order to adopt Kalman Filter Algorithm, a recursive process of two steps is involve; prediction and correction (update step). The current state is predicted from the previous state. The first process uses previous states to predict the current state. The second process uses the current measurement, such as human location, to correct the state. The recursive equations are as shown in equations 3 to 7. The derivations of these equations are beyond the scope of this paper. Interested persons should refer to [15].

A. Prediction

1) Predict the states ahead

$$\hat{x}_t^- = Ax_{t-1} + Bu_t \tag{3}$$

2) Predict the error covariance

$$P_t{}^- = AP_{t-1}A^T + Q \tag{4}$$

B. Correction (Measurement update)

1) Compute the Kalman Gain

$$K = P_t{}^- H^T (HP_t{}^- H^T + R)^{-1} \tag{5}$$

2) Update estimate with measurements z_t

$$\hat{x}_t = \hat{x}_t^- + K(z_t - H\hat{x}_t^-) \tag{6}$$

3) Update covariance error

$$P_t = (I - KH)P_t^- \tag{7}$$

$Where$, P is the prediction error covariance. Q is the process noise covariance, R is the measurement error covariance and K is the Kalman gain, A, B, H are the model matrices.

V. Tracking System Modeling

Human is assumed to be walking with constant velocity so the human current state vector needs to be predicted from the previous state (location). Bounding boxes are extracted from each frame and association of detection result to the tracked human in two consecutive frames are done using Hungarian technique [16] of assignment problem. Finally, the associated data serves as input to the Kalman Filter for parameters estimation.

A. Modeling State or Process equation

The human tracking problem from video can be modeled using Newton's dynamic equation.

$$x_t = x_{t-1} + v_x(\Delta t) + \frac{1}{2}a(\Delta t)^2 \tag{8}$$

$$y_t = y_{t-1} + v_y(\Delta t) + \frac{1}{2}a(\Delta t)^2 \tag{9}$$

Equation 8 is the distance covered in the x direction while equation 9 is the distance covered in the y direction. Δt is the discrete time interval. In this model, Δt will be the frame (sequence) number, v is the velocity at which human is moving in the video, a is the acceleration.

The state variables are the Cartesian coordinates in both x and y direction, velocity in both x and y direction, length, l

and width, w of the bounding box of the person of interest. Equation 8 and 9 can be re-written in vectors from in terms of this information available to us.

$$\begin{bmatrix} x \\ y \\ l \\ w \\ v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & 0 & 0 & \Delta t \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ l \\ w \\ v_x \\ v_y \end{bmatrix} + \frac{1}{2}a(\Delta t)^2 \quad (10)$$

$\Delta t = 1$, and acceleration a which is the rate of change of velocity is equal to zero (0).

$$x_t = Ax_{t-1} + w_{t-1} \quad (11)$$

$Where\ x_t$ is a state vector, w_{t-1} is white Gaussian noise which is independent and identically distributed with mean zero and Variance Q, that needs to be determined. A is state transition Matrix with value

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

B. Modeling the measurement equation

Measurement equations can be modeled in terms of

$$z_t = Hx_t + v_t \quad (12)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

C. Parallel Implementation

The NVIDIA GEFORCE is used for this work which uses the Compute Unified Device Architecture (CUDA). The advantage of CUDA architecture is that, Shared memory of each multiprocessor are exposed and this makes main memory easily accessible. Each function launched from the CPU to the GPU is called Kernel which has parameters that indicate the number of parallel section as well number of threads required for each parallel section. A thread block is threads that are working together in a section.

As the first step in the GPU implementation required the image data be copied from the CPU to the GPU main memory, the video data is transferred to the GPU. The GPU computes the HOG features and Haar features for human detection. Once human detection is done, the GPU invokes threads for each detected human hence Kalman object is instantiated for each new human detected and computes the estimation for the existing tracked human. The proposed algorithm is based on Kalman Filtering and this automatically create, initialize tracks and then assign unique identification (ID) for each new person detected for τ consecutive frames. τ is chosen to reduce false positive detection. In the experimental setup, $\tau = 5$. The issue of multiple human target management is addressed using an assignment problem. The detected object in the previous frames needs to be associated with the object detected in the current frame, this can include variable number of tracked object, birth of new tracks and death of an existing track. Data association of detected human in the current frame are matched according to the Hungarian technique of assignment problem [16]. The corrected tracked from the Kalman are sent back to the CPU to be compared with the grand truth values.

VI. Experiment and Results

Two series of experiments were carried out to evaluate the system. In the first experiment, we evaluated the detection module using only image dataset. after that, we evaluated the tracking model on video dataset. Both experiments were conducted using Intel i5 core duo CPU and each core frequency is 2.50 GHz. System memory of 10GB and NVDIA getforce 610M. MATLAB 2012a for Microsoft windows was used for the programming development.

A. Detection Evaluation

The Full body training and testing dataset comes from INRIA dataset [17], which is made up of 1000 positive (human) images and 500 negatives (non-human) images. 10-fold cross validation was used to train the model performance. To access the performance of the detection algorithm, standard performance measurement used in computer vision and image retrievals which are precision, recall and F score were used on the test dataset. The detail of the people detection is reported in [13]. Graph of precision against recall of the testing data is as shown in figure 2.

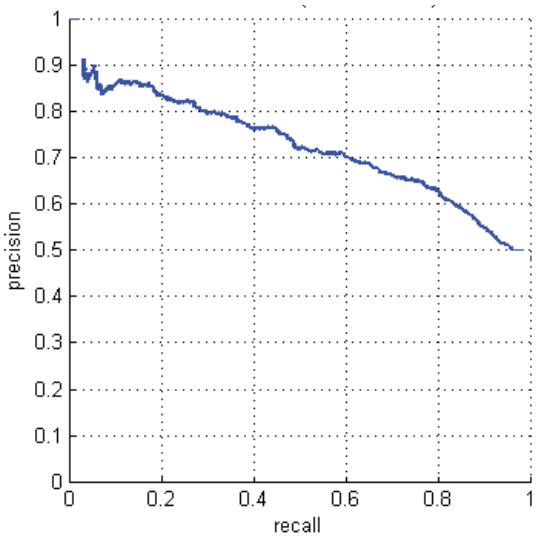

Fig. 2. Precision recall graph for Human detection

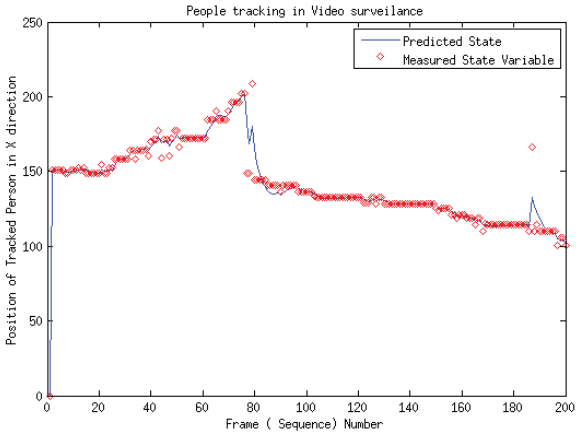

Fig. 3. Human tracking with Kalman Filter model

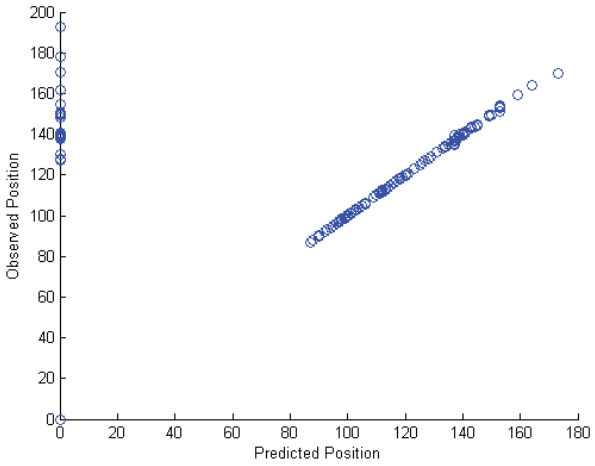

Fig. 4. Graph of Measured data against the predicted

B. Tracking Result

The model is evaluated on the video dataset which contain 1019 frames obtained from [18], UTP dataset (Computer Laboratories' surveillances) and $viptrain.avi$ video (650 frames) shipped with computer vision system toolbox. Kalman Filter is initialized when the first human detection is observed from the detection modules and the first detection serves as the initial state variables. Setting the initial state covariances was a problem but it was assumed that it is 100. Error covariance indicates the level of confidence we have on the data, therefore, setting it to be large value shows the uncertainty in the data. Error covariance will eventually converge to the actual value in the long run so it does not matter, whatever the initial value we set the covariance to. The tracking position in x direction is as shown in Figure 3. From the figure, the red diamond is the measured location while the blue lines is the Kalman filter estimated location. The step shown in Figure 3 is resulted from occlusion and disappearance of the tracked person. Figures 6 and 7 shows the tracked persons on some selected frames from the video.

Linearity assumption of Kalman Filter was tested on the dataset, observed data was plotted against the estimated data in the first component of the state variables. Result from the figure 4 supports the model as it is expected, most of the data points lie almost on a straight line at an angle of 45 degrees.

We further checked for dependency within the measurement error, by plotting the tracking error data. Figure 5 supports the randomness and independency of the tracking errors. This further supports the normality assumption of the state variables. The two spikes in the graph indicate a missing measured value due to occlusion. However the Kalman was still able to estimate and predicate the assumed location of the tracked person.

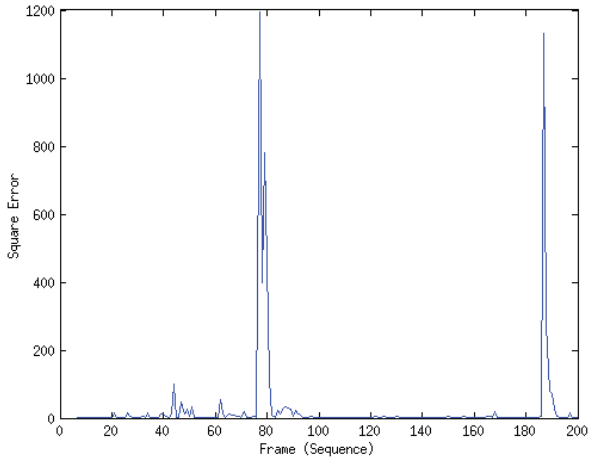

Fig. 5. Difference between measured value and the predicted position in the X- direction

Fig. 6. Tracking of single Person among many people. The video dataset is obtained from [18]

VII. Conclusion

Tracking is an integral part of most surveillances. This research work proposed a novel real time multiple human target tracker based on combined features; Histogram of Oriented Gradient and Haar Features for the human detection and

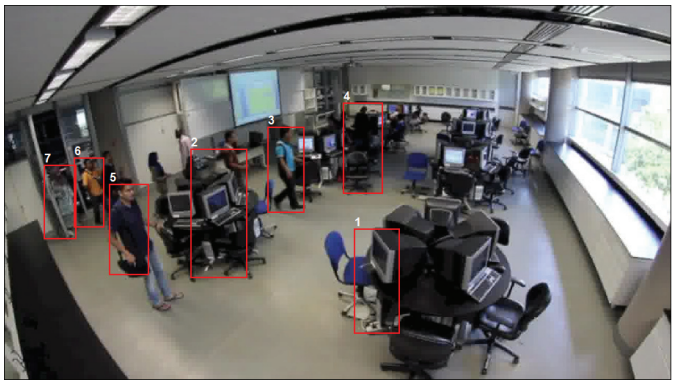

Fig. 7. A frame image of Multi Person Tracking with unique ID assigned to each person when first detected. The ID is associated with the person throughout the lifespan of the person.

utilizes Kalman Filtering applied on the parallel Architecture, GPU to perform multiple target tracking. Kalman Object are automatically instantiated once a new object are detected hence no need to limit the maximum number of targets a priori. This work demonstrates the modeling and application of Kalman Filter for human tracking using the result obtained from the detection module. The method when evaluated on the publicly available dataset for tracking gives a very good result.

REFERENCES

[1] M. Shah, O. Javed, and K. Shafique, "Automated visual surveillance in realistic scenarios," *Multimedia, IEEE*, vol. 14, no. 1, pp. 30–39, 2007.

[2] R. D. Lascio, P. Foggia, G. Percannella, A. Saggese, and M. Vento, "A real time algorithm for people tracking using contextual reasoning," *Computer Vision and Image Understanding*, 2013.

[3] V. H. Diaz-Ramirez, V. Contreras, V. Kober, and K. Picos, "Real-time tracking of multiple objects using adaptive correlation filters with complex constraints," *Optics Communications*, vol. 309, pp. 265–278, 2013.

[4] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, "W 4 s: A real-time system for detecting and tracking people in 2 1/2d," in *Computer VisionECCV'98*. Springer, 1998, pp. 877–892.

[6] A. Behrend, G. Schüller, and M. Wieneke, "Efficient tracking of moving objects using a relational database," *Information Systems*, 2012.

[7] F. Poiesi, R. Mazzon, and A. Cavallaro, "Multi-target tracking on confidence maps: An application to people tracking," *Computer Vision and Image Understanding*, 2012.

[8] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 9, pp. 1820–1833, 2011.

[9] H. Medeiros, G. Holguín, P. J. Shin, and J. Park, "A parallel histogram-based particle filter for object tracking on simd-based smart cameras," *Computer Vision and Image Understanding*, vol. 114, no. 11, pp. 1264–1272, 2010.

[10] J. Saboune and R. Laganiere, "People detection and tracking using the explorative particle filtering," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1298–1305.

[11] A. M. Abdel Tawab, M. Abdelhalim, and S.-D. Habib, "Efficient multi-feature pso for fast gray level object-tracking," *Applied Soft Computing*, vol. 14, pp. 317–337, 2014.

[12] S. Tomar, "Converting video formats with ffmpeg," *Linux Journal*, vol. 2006, no. 146, p. 10, 2006.

[13] A.-L. Yussiff, S.-P. Yong, and B. B. Baharudin, "People detection enrichment for abnormal human activity detection," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 8, pp. 632–640, 2013.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[15] M. S. Grewal and A. P. Andrews, "Kalman filtering: theory and practice using matlab," *John Wiley & Sons, Baltimore, MD*, vol. 2, p. 35, 2001.

[16] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[18] D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, Oct 2010, pp. 772–777.