SICE Annual Conference 2014
September 9-12, 2014, Hokkaido University, Sapporo, Japan

Kalman Filter Based Multiple Objects Detection-Tracking Algorithm Robust to Occlusion

Jong-Min Jeong[1†], Tae-Sung Yoon[2], and Jin-Bae Park[1]

[1]Department of Electrical and Electronic Engineering, Yonsei University, Seoul, 120-749, Korea
(Tel : +82-2-2123-2773; E-mail: {jeong6560, jbpark}@yonsei.ac.kr)
[2]Department of Electrical Engineering, Changwon National University, Changwon, 641-773, Korea
(Tel : +82-55-213-3633; E-mail: tsyoon@changwon.ac.kr)

Abstract: Visual target tracking is one of the major fields in computer vision system. Object tracking has many practical applications such as automated surveillance system, military guidance, traffic management system, fault detection system, artificial intelligence and robot vision system. But it is difficult to track objects with image sensor. Especially, multiple objects tracking is harder than single object tracking. This paper proposes multiple objects tracking algorithm based on the Kalman filter. Our algorithm uses the Kalman filter as many as the number of moving objects in the image frame. If many moving objects exist in the image, however, we obtain multiple measurements. Therefore, precise data association is necessary in order to track multiple objects correctly. Another problem of multiple objects tracking is occlusion that causes merge and split. For solving these problems, this paper defines the cost function using some factors. Experiments using Matlab show that the performance of the proposed algorithm is appropriate for multiple objects tracking in real-time.

Keywords: Kalman filter, Multiple objects tracking, Data association.

1. INTRODUCTION

Object tracking is one of the important fields in computer vision. The tracking of moving objects has many issues in video stream such as automated surveillance, military guidance, traffic management system, robot vision and artificial intelligence [1]. Recently, a variety of algorithms available for moving single object have been proposed, and most of them showed successful performance. In the real world, however, most cases are the situation where several objects exist. Contrary to single object tracking, there are many problems in multiple objects tracking. One of the important problems is matching between targets and observations, data association, from frame to frame in a video sequence. In order to perform multiple objects tracking successfully, matching between targets and observations should be performed correctly from frame to frame. However it is hard task in computer vision fields. Especially, if several objects occlude each other, it is more difficult to match correctly. Many researchers have done a lot of works about this topic [2,3].

To solve these problems, Shiloh et al. [4], Chang et al. [5], and Dockstader [6] overcame occlusion in multiple objects tracking using multiple camera. Wu [7] proposed a dynamic Bayesian network which uses an extra hidden process in order to handle a partial occlusion. K. HyunBok et al. [8] proposed a particular object tracking based on condensation filter. But this method did not consider occlusion in the environment where multiple objects exist. Tao Yang et al. [9] used feature correspondence for occlusion handling in dynamic scenes. Though this algorithm is robust and accurate, it has highly computation cost. L. Ido et al. [10] proposed mean shift based object tracking. It is computationally efficient, but sensitive to background and occlusion. Although various algorithms have been proposed for successful multiple object tracking, the problem of multiple object tracking in dynamic scene is still far from being completely solved.

This paper considers the problem of simultaneously tracking one or more objects in video sequence. In particular, our paper focuses on the cases where several objects occlude each other, either partially or completely. To deal with multiple objects tracking in dynamic scenes, we proposed a Kalman filter based tracking algorithm. The Kalman filter provides a robust object tracking framework under uncertain environments. But since we should handle occlusions, some features are used. Through the cost function that we define, we solve the data association problems effectively. Experimental results show the proposed algorithm is able to do efficient and robust multiple objects tracking with occlusion problems in dynamic scenes.

This paper is organized as follows: In section 2, overview of the proposed algorithm is explained briefly. Section 3 consists of 4 subsections and describes the proposed algorithm in detail. Then in section 4, experimental results and some discussion are presented. Finally, section 5 includes the conclusion of our paper.

2. SYSTEM OVERVIEW

The goal of our work is to track the multiple objects in the dynamic scenes. In order to achieve this purpose, this paper proposes the Kalman filter based multiple objects tracking algorithm. First, a background subtraction and motion information is used for detecting multiple moving objects. Then, we find out the number of objects in the frame. Second, Kalman filters are created as many as the number of detected objects. In order to accurate tracking, however, we should select a true measurement for each standard Kalman filter. So, our proposed algorithm determines which measurement

† Jong-Min Jeong is the presenter of this paper.

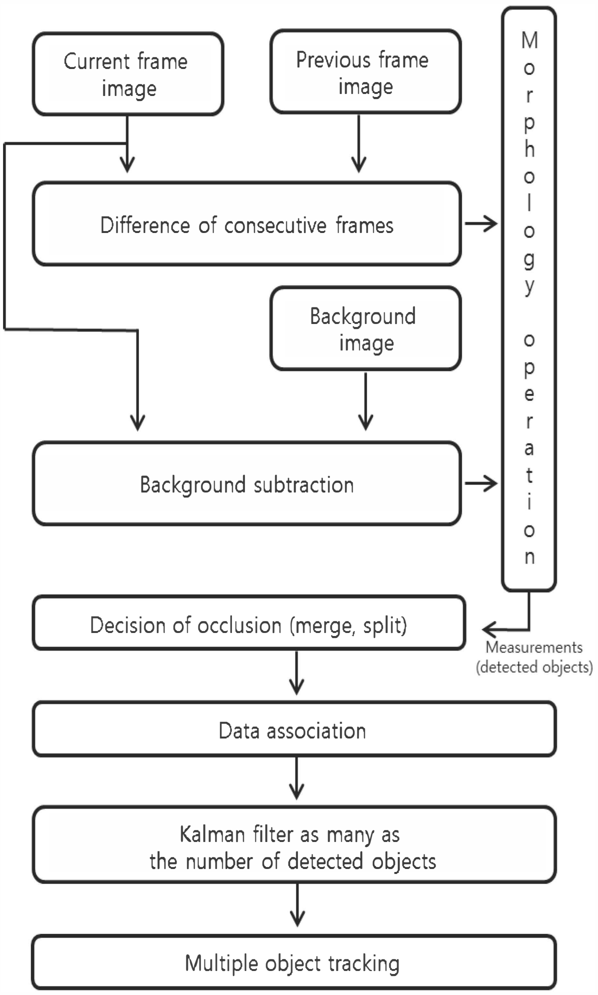

Fig. 1 The flow chart of the proposed algorithm

corresponds to each Kalman filter among several measurements using cost function that consist of some features. Also, when multiple objects are present, there are object occlusion and split problems. Proposed algorithm determines whether merge or split is occurred or not by applying some features. Figure 1 shows the flow chart of the proposed algorithm.

3. PROPOSED MULTIPLE OBJECTS TRACKING ALGORITHM

3.1 Moving objects detection

The proposed algorithm uses some information of multiple objects in the tracking process. To obtain the information such as positions and the number of pixels that objects occupy, we should detect multiple objects in the frame. So, we use several algorithms in this process. First, difference of consecutive frames is used to detect the change area of frames. Since moving objects are considered as targets to be tracked, it is operated well.

$$\mathrm{FD}_t(x,y) = \begin{cases} 0 & if\ |I_t(x,y) - I_{t-1}(x,y)| < \tau_{FD} \\ 1 & if\ |I_t(x,y) - I_{t-1}(x,y)| \geq \tau_{FD} \end{cases} \quad (1)$$

Where $I_t(x,y)$ is the gray image at frame t. τ_{FD} is the the threshold value. If the change of objects is small in certain frame, however, the algorithm may not detect all of moving objects. Therefore, our algorithm utilizes also background subtraction. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. In proposed algorithm, assume that background is already updated every 10 frames before tracking process. It is expressed as follows:

$$BS_t(x,y) = \begin{cases} 0 & if\ |I_t(x,y) - B(x,y)| < \tau_{BS} \\ 1 & if\ |I_t(x,y) - B(x,y)| \geq \tau_{BS} \end{cases} \quad (2)$$

Where $B(x,y)$ is the background that is already obtained. τ_{BS} is the threshold value. Through Eqs (1) and (2), we can detect moving objects in videos. The $BS_t(x,y)$ has accurate information more than $FD_t(x,y)$ relatively, however, since $FD_t(x,y)$ has many noises caused by lightness. For this reason, bit operation is used as follows:

$$\mathrm{BM}_t(x,y) = BS_t(x,y) \cup (BS_t(x,y) \cap FD_t(x,y)) \quad (3)$$

Where $\cup$ is the OR operation, $\cap$ is the AND operation. Finally, morphology operations are applied in order to get precise information. Especially, dilation and erosion are used. Dilation operation adds layers to objects, so it can enlarge small detected objects. Erosion operation peels off layers from objects, so it can remove extraneous pixels caused by dilation operation. These morphology operations are applied as follows:

$$\mathrm{ME}_t(x,y) = Erode\left(Dilate\left(BM_t(x,y)\right)\right) \quad (4)$$

Figure 2 shows flow chart that describes detection

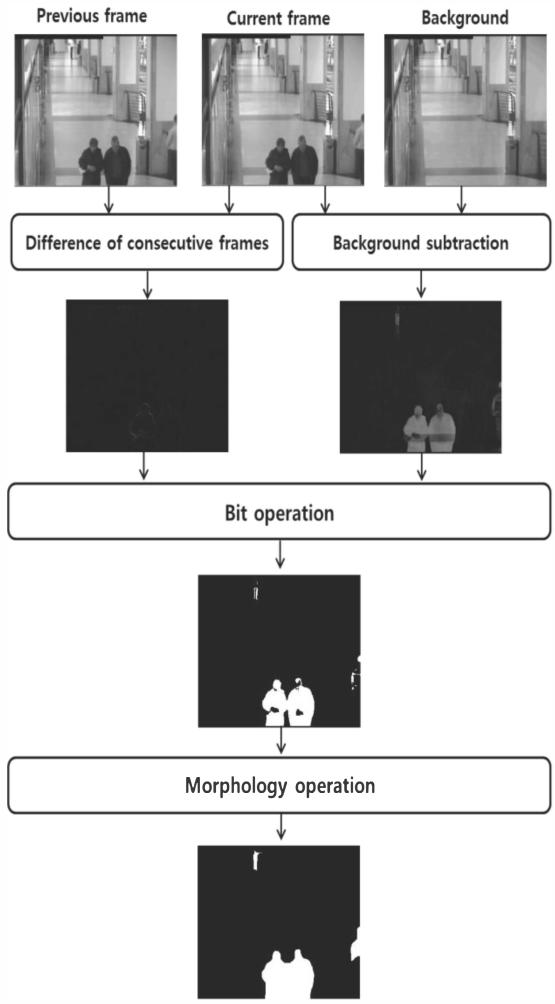

Fig. 2 Objects detection algorithm

algorithm in detail. Finally, we determine whether detected pixels are the part of the object to be tracked using the number of detected pixels. Usually, a lot of pixels are occupied by the object. In contrast, noise is composed of the small number of pixels. Therefore, we can verify the detected pixels using the number of detected pixels (pixel area). The verification process is as follows.

$$\begin{cases} Object & if\ A_i > \tau_A \\ Noise & Otherwise \end{cases} \tag{5}$$

Where A_i is the pixel area of i^{th} detected pixels' group and τ_A is the threshold value.

3.2 Kalman filter

The Kalman filter is an estimator that provides an efficient recursive method to estimate the state of a linear process, in a way that minimizes the mean of the squared error [11]. The Kalman filter is typically divided into two stages. One is the time update (prediction), the other is the measurement update (correction). Time update is to advance the state based on state equation until the next measurement is obtained. Measurement update is to incorporate the measurement from sensors based on measurement equation. In the proposed algorithm, the same number of the Kalman filter as objects are detected are deployed in order to estimate each object's state. Each Kalman filter estimates the position, (x, y) in the frame, of each object to be tracked. Since the standard Kalman filter uses one correct measurement, data association should be considered to classify true measurement and false measurements. Each Kalman filter is configured as follows:

$$x_k = Ax_{k-1} + w_k \tag{6}$$

$$z_k = Hx_k + v_k \tag{7}$$

Where

$$\mathrm{x} = \begin{bmatrix} p_x & p_y & v_x & v_y \end{bmatrix}^T$$

$$\mathrm{A} = \begin{bmatrix} 1 & 0 & \mathrm{T} & 0 \\ 0 & 1 & 0 & \mathrm{T} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathrm{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

p_x, p_y represent the center position of x-axis, the center position of y-axis, and v_x, v_y are the velocity of x-axis and y-axis. Matrix A represents the transition matrix, matrix H is the measurement matrix, and T is the time interval between two adjacent frames. w_k, v_k are the Gaussian noises with the error covariances Q_k and R_k. Process of the Kalman filter is as follows.

1) Time update of the state estimate

$$x_{k|k-1} = Ax_{k-1|k-1} \tag{8}$$

2) Predicted measurement

$$z_{k|k-1} = Hx_{k|k-1} \tag{9}$$

3) Time update of the state error covariance

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q_k \tag{10}$$

4) Data association process

In order to deploy the Kalman filter to the tracking of multiple objects' location in an environment where many measurements exist, matching between object and measurement should be performed correctly. So, several factors are used for matching process in the proposed algorithm. It will be described in detail in section 3.3 and 3.4.

5) Kalman gain

$$K_k = P_{k|k-1}H^T\left(HP_{k|k-1}H^T + R\right)^{-1} \tag{11}$$

A Kalman gain depends on the accuracy of a measurement. If an accuracy of the measurement is high, the Kalman gain has high value. Otherwise, the Kalman gain has relatively low value.

6) Measurement update of the state error covariance

$$P_{k|k} = (I - KH)P_{k|k-1} \tag{12}$$

Where I is a 4 by 4 unit matrix.

7) Measurement update of the state estimate

$$x_{k|k} = x_{k|k-1} + K_k(z_k - z_{k|k-1}) \tag{13}$$

3.3 Decision of occlusion

In multiple objects tracking, occlusion is one of the important problems. Occlusion essentially includes merge and split problems. Combining two or more objects into one is called merge problem. On the contrary, separating from merging objects is called split problem. Merge and split are illustrated in Figure 3. This section covers how to determine which case is occurred. Proposed algorithm can distinguish it using ratio variation.

1) Merge problem

In order to check whether the merge problem happens or not, proposed algorithm uses the fact that the ratio of merged object is different from not merged object. The ratio is defined as follows.

$$R_i = \frac{Height_i}{Width_i} \tag{14}$$

Merge condition is expressed as follows:

$$R_k^i > \tau_{ratioUp}, \ \ i = 1, \dots, m \tag{15}$$

$$R_k^i < \tau_{ratioDown}, \ \ i = 1, \dots, m \tag{16}$$

Where R_k^i is the ratio of the object obtained from i^{th} measurement at frame k. $\tau_{ratioUp}$ is the upper threshold value about the aspect ratio and $\tau_{ratioDown}$ is the lower threshold value about the aspect ratio, m is the number of detected objects at frame k.

2) Split problem

Split problem always is occurred after merge problem is happened. In the split problem, we also use the fact that the ratio of split object is similar to it of the single object. Therefore, the condition of split event is defined as follows:

$$\tau_{ratioDown} < R_k^i < \tau_{ratioUp}, \ \ i = 1, \dots, m \tag{17}$$

In order to use the Kalman filter in multiple object tracking, data association that is matching process between targets and measurements is necessary. Its explanation is described in detail in section 3.4.

3.4 Data association

This paper deals with multiple objects tracking. In such an environment that has many objects, we can obtain multiple measurements through detection algorithm. In order to track objects correctly, however, data association is necessary. But it is difficult to determine what measurement is legitimate to each

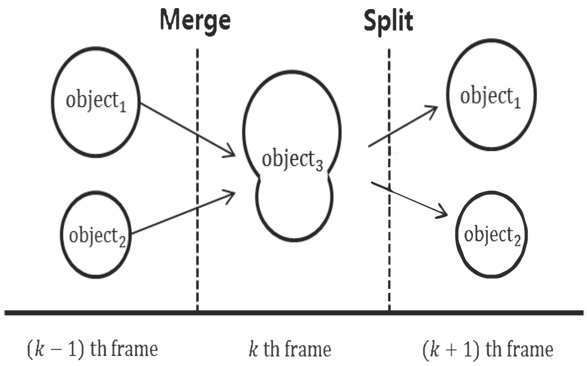

Fig. 3 Illustration for merge and split

object. So, to classify the correct measurement corresponding to each object from many measurements, we use the distances between the estimated positions and the measurements and area variations.

1) Distance

As one of the determining factor, distance is used between the latest positions of targets to be tracked and the positions of the obtained measurements. If it has smaller value, the probability that corresponding measurement is true is higher.

$$D_k(i,j) = \frac{\left|\sqrt{\left(p_{x_j}^{k-} - z_{x_i}^{k}\right)^2 + \left(p_{y_j}^{k-} - z_{y_i}^{k}\right)^2}\right|}{max\left|\left(p_{x_j}^{k-} - z_{x_i}^{k}\right)^2 + \left(p_{y_j}^{k-} - z_{y_i}^{k}\right)^2\right|}, \quad \begin{matrix} i = 1, \dots, m \\ j = 1, \dots, n \end{matrix} \qquad (18)$$

Where $p_{x_j}^{k-}$ and $p_{y_j}^{k-}$ is the center position of x and y obtained from the $x_{k|k-1}$ in the j^{th} Kalman filter. ${z_x}_i^k$ and ${z_y}_i^k$ is the center position of x and y obtained from i^{th} measurement at frame k. m is the number of detected measurements at frame k and n is the number of targets in frame k-1.

2) Area

Another factor is the area that objects occupy. This factor is defined as follows in order to data association.

$$A_k(i,j) = \frac{\left|A_k^i - A_{k-1}^j\right|}{max\left|A_k^i - A_{k-1}^j\right|}, \quad \begin{matrix} i = 1, \dots, m \\ j = 1, \dots, n \end{matrix} \qquad (19)$$

The smaller this value is, the higher the probability of the corresponding measurement being true is.

By combining of Eqs. (18) – (19), we define the cost function. If merge or split problem are occurred, cost function depends on only distance. Since area variation changes abruptly at that time, it cannot be used as cost function for data association.

$$C_k(i,j) = D_k(i,j) \qquad (20)$$

Eqs (20) is the cost function for data association when merge or split is occurred.

Otherwise, if merge or split problem is not occurred, the cost function consists of distance and area as follows.

$$C_k(i,j) = \alpha D_k(i,j) + \beta A_k(i,j) \qquad (21)$$

Where $\alpha + \beta = 1$, this parameter can be set experimentally. We can assume that i^{th} measurement correspond to j^{th} object if $C_k(i,j)$ has the smallest value.

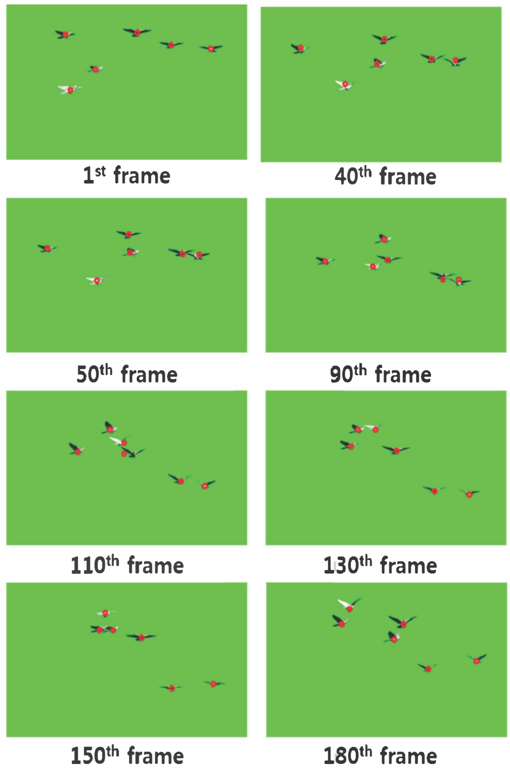

Fig. 4 Tracking results of the proposed algorithm with the flock of similar bird video

4. EXPERIMENT RESULTS

In this part, we adopt two videos to test the performance of the proposed algorithm. The experiment is implemented using Matlab 2012a on the Microsoft windows 8 and Intel® Pentium® CPU G620 with 8G RAM. The first video displays the flock of similar birds moving. This video consists of 350 frames and each frame has 360×640 pixels. Frame rate is 30frames/s. Figure 4 shows the tracking results of the proposed algorithm about the first video. At the 50$^{\text{th}}$ frame, two birds are merged each other and occlusion problem begins to occur. In the rest frames, occlusion problems continuously are occurred. Although there are many difficult situations, the proposed algorithm has a good performance. The second video displays walking persons in the hall. It consists of 485 frames, and each frame has 288 × 384 pixels. This video also has occlusion circumstances. Even if occlusion happens at the 200$^{\text{th}}$ frame, the proposed can track objects correctly. Also proposed algorithm can check new appeared objects and disappeared objects. At the 10$^{\text{th}}$ frame, two objects adjoin each other closely. But this algorithm can recognize them as two objects instead it recognizes as one object. According to these results, it is robust and efficient algorithm. Figure 5 shows the tracking results of the proposed algorithm about the second video.

Table 1 shows the average calculation time per 1 frame

Fig. 5 Tracking results of the walking people in the hall video

of each simulation. If the number of objects is large, the average calculation time increases. But there is no difficulty in operating the algorithm real-time since the calculation time is short enough. Considering the calculation time and the performance, it is expected that the proposed algorithm is useful to track multiple objects in the real-world.

Table 1 calculation time of each case

	First video	Second video
Time(s)	0.00236	0.00197

5. CONCLUSION

In this paper, we proposed the Kalman filter based multiple objects tracking algorithm. Background subtraction, morphology operation and difference of consecutive frames are used for multiple objects detection. Through this process, some information such as the positions and area of objects is obtained. In an environment where many objects exist, however, it is difficult to distinguish each measurement correctly. In order to solve these problems, two factors including distance and area are used. Also, occlusion related to merge and split is dealt with. The proposed method solved by setting the different cost function. Finally, through the experiment results, we showed that the proposed algorithm is suitable for real-time multiple tracking.

ACKNOWLEDGEMENT

This work has been supported by Institute of BioMed-IT, Energy-IT and Smart-IT Technology (BEST), a Brain Korea 21 plus program, Yonsei University

REFERENCES

[1] Y. Hanxuan, S. Ling, Z. Feng, . Liang, and S. Zhan, "Recent advances and trends in visual tracking: A review", *Neurocomputing*, Vol. 74, No. 18, pp. 3823-3831, 2011

[2] L. Xin, W. Kejun, W. Wei, and L. Yang, "A Multiple Object Tracking Method Using Kalman Filter", *IEEE International Conference on Information and Automation*, pp. 1862-1866, 2010.

[3] C. Srinivas, A. Hoogs, G. Brooksby, and H. Wensheng, "Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 666-673, 2006.

[4] S.L. Dockstader, and A.M. Tekalp, "Multiple Camera Tracking of Interacting and Occluded Human Motion", *Proceedings of the IEEE*, Vol. 89, Issue. 10, pp. 1441-1455, 2002.

[5] Ting-Hsun Chang, Shaogang Gong, and Eng-Jon Ong, "Tracking Multiple People under Occlusion Using Multiple Cameras", *In Proc. 11th British Machine Vision Conference*, 2000.

[6] S.L. Dockstader, and A.M. Tekalp, "Multiple Camera Fusion for Multi-Object Tracking", *In Proc. IEEE Workshop on Multi-Object Tracking,* pp. 95-102, 2001.

[7] Ying Wu, Ting Yu, and Gang Hua, "Tracking Appearances with Occlusions", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Vol. 1, pp. 789-795, 2003.

[8] K. HyungBok, and S. KweeBo, "A Particular Object Tracking in an Environment of Multiple Moving Objects", *International Conference on Control, Automation and Systems,* pp. 1053-1056, 2010.

[9] Tao Yang, Stan Z.Li, Quan Pan, and Jing Li, "Real-Time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes", *IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 970-975, 2005.
[10] L. Ido, L. Michael, and R. Ehud, “Mean Shift Tracking with Multiple Reference Color Histogram”, *Computer Vision and Image Understanding,* Vol. 114, No. 3, pp. 400-408, 2010.
[11] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Journal of Fluids Engineering*, Vol. 82, Issue. 1, pp. 35-45, 1960.