

Least Square Mean Optimization-Based Real Object Detection and Tracking

Sanjay Rao, Deepak Jhanwar, Diwakar Gautam
and Anand Choudhary

Abstract In this paper, automatic real-time object detection and tracking is implemented via means of Kalman filter in which the system output is actually tracking the input canceling out any variation due to input and output noises. This paper can be used to develop a surveillance system of static camera and robotic automation visual systems. Whenever a new object comes in camera frame, system uses the concepts of frame subtraction then threshold image by Otsu's method, and later Kalman filtering is being processed to estimate the next following coordinates of its movement. The work presented here is extended to work at video processing stage. And finally, least square mean optimization technique is used to evaluate the set of system parameters for perfect tracking of forthcoming new objects, and once that parameter is evaluated it be can used to execute tracking process perfectly.

Keywords Kalman filter · Object detection · Otsu's method · Least square mean optimization · Automatic visual systems

1 Introduction

The real-time object detection and tracking has been a great field of research since emergence of field of Computer Vision and Image Processing. Earlier, many great contributions had been done by various scholars in this field. The video surveillance

S. Rao (✉) · D. Jhanwar · D. Gautam · A. Choudhary
Department of ECE, Government Engineering College Ajmer,
N.H. 8 Badliya Circle, Ajmer, India
e-mail: sanjaym.rao2012@gmail.com

D. Jhanwar
e-mail: deepakjhanwar2001@gmail.com

D. Gautam
e-mail: gdiwakar1987448@gmail.com

systems can be classified under two broad categories: static camera systems and moving camera systems. The work presented here is majorly concentrated on static systems while the presented concepts can be extended to moving camera systems by timely varying their reference frames. Background subtraction technique has been used a lot in previous works. But as technology develops, the processing time of new algorithms continues to shrink, here is too proposed a new method that is fast in processing with a fine result. Object detection can be done by two methods: (a) automatic systems and (b) manual systems. The manual system requires some human interference to locate any figure on the foreign object [1, 2], while in automated systems once the parameters have got set it can detect new foreign object by itself. In modern systems, this system can be implemented too by using color or texture information [3] of foreign objects in the frame. In previous approach, there has been a use of reference coordinates in the system to identify new objects; these coordinates can be obtained by taking edges of fixed bodies in the reference frames that restrict to perform within a certain class of surveillance systems. In advancement to that, this paper works on single- or multi-object detection by using morphological operations in the field of signal processing. The work presented here can be subdivided into various sections. Section A deals with object detection via averaging out histogram differences. Section B works with thresholding which is implemented using Otsu's threshold [4], but different methods can be used too for the same. Once the object is extracted from background, then morphological operations are used to detect number of new objects. Section C works on Kalman filter to estimate the next coordinates of object motion [5].

2 Otsu's Threshold and Class Separability

Histogram is probability representation of different gray levels in a given plane of three-dimensional images. Thus, thresholding is subdividing the whole system into two parts: foreground object and background object. There must be clear-cut valley in the histogram to easily evaluate the threshold value of gray level to subdivide our images. But this may not be the case for always, as many a times noise degrades the deep valley. In such a case, Otsu's method can be used to extract object in the image. In this method, the procedure followed in such a way that for every possible value of threshold, histogram is subdivided into classes. Then, total image variance, within the class variance and between the class variance, is used to evaluate the most exact value of threshold. This test is basically a measure of class separability. The gray image is converted into binary image using global threshold technique of Otsu's threshold [1]. If input is RGB image, result of Otsu's threshold in each plane is concatenated.

3 Object Detection and Extraction

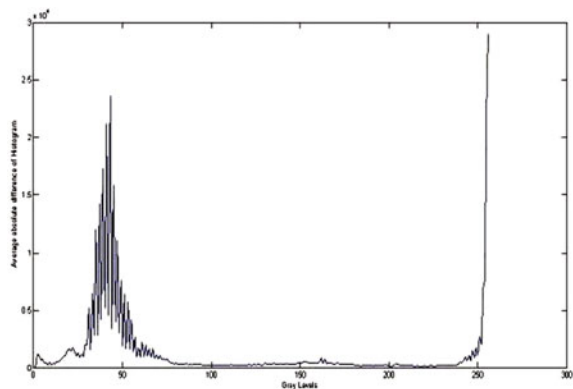
The histogram display of subsequent image sequences is frequency component, low-pass filtering will eliminate it, and result will be a more fine view of foreign labeling information of number of objects, which can be used as basic parameter to initiate Kalman filtering approach discussed later on. The detection of the foreign frames in consecutive frames is shown in Fig. 1, subtracted consecutively, and averaged out. The resultant peak at certain region of gray level indicates that the newcoming foreign object has gray scale in that particular range. The threshold image either from RGB plane or from single gray plane is first of all converted into a binary image. Then, morphological operations of image labeling are used to detect the number of objects in the image; prior to this operation, the extracted object is filtered with a low-pass filter to eliminate any noise in the image, as noise is high.

4 Discrete Kalman Filter

In order to track the moving objects, discrete version of Kalman filter is used. Kalman filter is basically estimator to follow system response irrespective of input or output and any other inherent system noises discussed later on. The detection of the foreign frames in consecutive frames is shown in the figure.

The filter predicts the next movement of object depending upon the parameters of previous and present state. Then, it measures variation of the observed value from predicted one. Thus, Kalman filter can be understood to work in two steps: (a) predictive stage and (b) measurement or correction stage analysis. As a paper is mostly concentrated on motion analysis, this paper is based on two types of error estimates: priory state estimates and posterior state estimates. Priory estimate is the prediction of next state parameters using the information of previous state, before the actual process is going to be occurred. Posterior estimate refers to state

Fig. 1 Object detection by averaging out histogram differences



estimation once the actual process or measurement has got completed. The complete module of the proposed technique is shown in the figure below:

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

$$x_k = \left[x, y, \frac{dx}{dt}, \frac{dy}{dt} \right]' \tag{2}$$

where A is state transition matrix relating the present state of process to its previous state and x_k represents system state at k th instance.

The state of system is represented via four important state variables: (a) x coordinate, (b) y coordinate, (c) directional velocity in x coordinate, and (d) directional velocity in y coordinate; thus, A can be represented in matrix notation as follows:

$$A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where dt represents the time duration between consecutive frames. The measurement process of system response can be represented as follows:

The system inherent and measurement noise parameters are w and v , respectively. These are white noise independent of each other possessing normal probability distribution $p(w) N(0, Q)$ and $p(v) N(0, R)$. Q is variance of process inherent noise and R is that of measurement noise (error in measurement). Z_k presents the measurement at k th instance of system process. As it seems not an easy task to measure state variables, so as to convert these state variables into measurable quantities, H matrix is denoted as follows:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{4}$$

In case if the effect of noises is very small, then residual signal is too small, so x_k is equal to \bar{x}_k , while this may not happen under significant effect of noises. In order to work under such circumstances, a priory state of x_k needs to be utilized represented as $\hat{x}_{\bar{k}}$. Thus, in priory state, the system and measurement equations during the priory state of x_k are expressed as $\hat{x}_{\bar{k}}$, and similarly whole system equations are represented as follows:

$$\hat{x}_{\bar{k}} = A\hat{x}_{\bar{k}-1} \tag{6}$$

$$\hat{Z}_{\bar{k}} = H\hat{x}_{\bar{k}} \tag{7}$$

Similarly, the posterior state represents the ideal value of state variable neglecting any noise responses, x_k is represented as $\hat{x}_{\bar{k}}$ and, respectively, others. Then, the priory and posterior error estimates are expressed as follows:

$$e_{\bar{k}} = x_k - \hat{x}_{\bar{k}} \tag{8}$$

$$e_k = x_k - \hat{x}_{\bar{k}} \tag{9}$$

Associated with these errors is a mean square error, or error variance can be related as

$$\begin{aligned} P_{\bar{k}} &= E[e_{\bar{k}}e_{\bar{k}}^T] \\ P_k &= E[e_k e_k^T] \end{aligned} \tag{10}$$

As the system at a given state depends upon elements of previous state, it is vital to initialize system with some default parameters. Thus, initialization recommends null vector as starting state variables and first state covariance matrix as zero matrix. From Fig. 2, it is easy to notify that at every iteration of filtering, Kalman gain K_k varies and also error covariances Q and R (measurement noise variance). Finally, the combined function of updating and measurement can be regrouped as follows:

$$P_{\bar{k}} = AP_{k-1} + Q \tag{11}$$

$$K_k = P_{\bar{k}}H^T (HP_{\bar{k}}H^T + R)^{-1} \tag{12}$$

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k(Z_k - H\hat{x}_{\bar{k}}) \tag{13}$$

$$P_k = (1 - K_kH)P_{\bar{k}} \tag{14}$$

Fig. 2 Discrete Kalman filter model

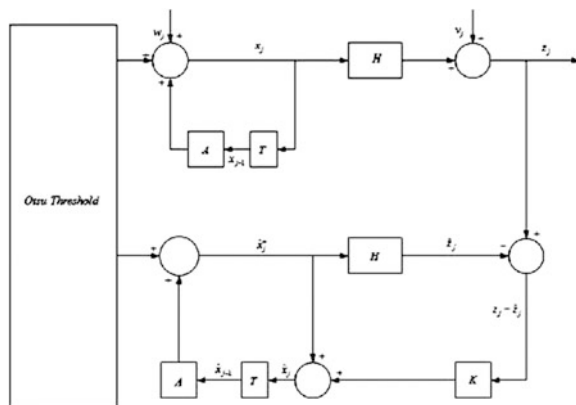
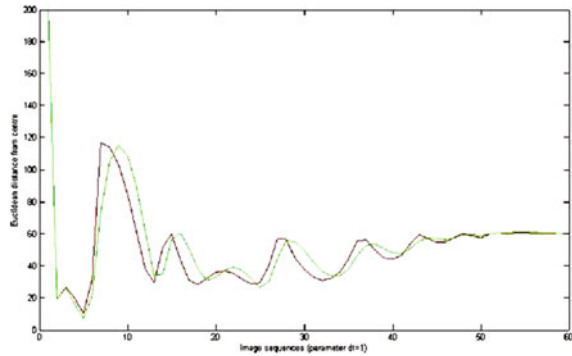


Fig. 3 Object tracking for lower values of frame occurrence rate $dt = 1$



The above equations are iterated for consecutive frames to exactly track up coordinates of moving objects. The noise variance R can be evaluated from static frames representing the noise variance executed in the measurement of the object coordinates. Thus, conclusions of Kalman filtering process can be described as follows (Fig. 3):

- (a) If the a priori error is very small, K_k is correspondingly very small, so our correction is also very small. In other words, we will ignore the current measurement and simply use past estimates to develop new estimate.
- (b) If the a priori error is very large, then in effect it tells to throw out the priority estimate and use the current (measured) value of the output to estimate the state.

5 Optimization of Frame Occurrence Timing

In actual practice, the frame extraction rate from general video sequence is 25 frames per second. But in this processing environment, it is very difficult to exactly define the time difference between consecutive frames due to above-specified conclusions of Kalman filtering process. The least mean square optimization or minimum deviation technique is implemented to evaluate the average frame occurrence rate, so that the system can track exactly moving objects, i.e., there must be a complete overlap among object centroid and tracking coordinates. One may use different notation of distance measurement such as absolute difference or Euclidean measurement to evaluate frame timing parameter by calculating mean square error between estimated measurement and actual measurement. Optimization result is figured as below.

Thus, it can be said that at higher values of frame occurrence rate, there tends to be minimum deviation between estimated and observed values.

6 Results and Conclusion

After applying the above-proposed technique on a practical video file, it seems very deterministic to detect new foreign object in image sequences or in a video file, the Red Cross represents the centroid of foreign object, and Green Square represents its Kalman filter estimated coordinates of its position at current state of system progress. The important point to be noted here is in Figs. 4 and 5, the variation in system parameter of frame occurrence timing, and Fig. 4 represents the result of unoptimized parameter and shows no overlap means moving object is not tracked perfectly while optimized result is shown in Fig. 5 where dt parameter is evaluated from whole system performance showing most perfect object tracking. Once this parameter is evaluated, it can be wholly utilized for the best result for any video or image sequences for this system. From Fig. 6, it is clearly visible that our approximation precisely matches the actual position of the object, giving us the authentication of the algorithm.

Fig. 4 Object tracking for optimized frame occurrence rate $dt = 100$

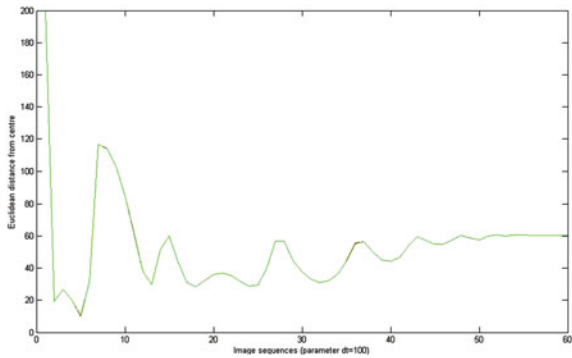


Fig. 5 Optimization of frame occurrence timing

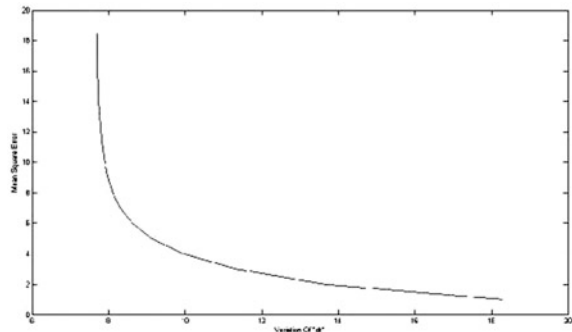


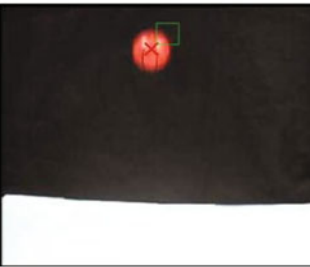







Image sequence	Frame rate(dt =1)	Frame rate (dt=100)
1.		
2.		
3.		
4.		

Fig. 6 Image sequences of real-time object tracking

References

1. Rother C, Kolmogrov V, Blake A. Grab cut interactive foreground extraction using iterated graph cuts. In: Proceedings of the SIGGRAPH, 2004.
2. Tan K-H, Ahuja N. Selecting objects with freehand sketches. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
3. Ning J, Zhang L, Zhang D, Wu C. Robust object tracking using joint color-texture histogram. *Int J Pattern Recogn Artif Intell.* 2009;23:1245–63.
4. Mehta M, Goyal C. Real time object detection and tracking: histogram matching and Kalman filter approach. *IEEE Conf.* 2010;5:796–801.
5. Welch G, Bishop G. An introduction to the Kalman filter. In: Presented at Annual Conference on Computer Graphics & Interactive Techniques, ACM SIGGRAPH, 2001, p. 201–4, 2001.