Human Detection and Tracking using Image Segmentation and Kalman Filter

Thombre D.V.
[1]Terna College of Engg Navi Mumbai.
dvthombre@rediffmail.com

Nirmal J H
K J Somaiya college of Engg, Mumbai,
asmita_nirmal@yahoo.com

Lekha Das
K J Somaiya college of Engg, Mumbai,
lekha2005@gmail.com

***Abstract*—Human detection and tracking is one of the most popular areas of video processing and the essential requirement of any surveillance system. In this paper we have used image segmentation technique for human detection and kalman filter with two dimension constant velocity model for human tracking. The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean square error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. This method tracks individual pedestrians as they pass through the field of vision of the camera, and uses vision algorithms to classify the motion and activities of each pedestrian. The tracking is accomplished through the development of a position and velocity path characteristic for each pedestrian using a Kalman filter. With this information, the system can bring the incident to the attention of human security personnel. In future applications, this system could alert authorities if a pedestrian displays suspicious behavior such as: entering a secure area, running or moving erratically, loitering or moving against traffic, or dropping a bag or other items**

Keywords- Human tracking, background subtraction, kalman filter

I. INTRODUCTION

The problem of using vision to track and understand the behavior of human beings is a very important one. It has applications in the areas of human-computer interaction, user interface design, robot learning, and surveillance, among others. At its highest level, this problem addresses recognizing human behavior and understanding intent and motive from observations alone. This is a difficult task, even for humans to perform, and misinterpretations are common. In the area of surveillance, automated systems to observe pedestrian traffic areas and detect dangerous action are becoming important. Many such areas currently have surveillance cameras in place, however, all of the image understanding and risk detection is left to human security personnel. This type of observation task is not well suited to humans, as it requires careful concentration over long periods of time. Therefore, there is clear motivation to develop automated intelligent vision-based monitoring systems that can aid a human user in the process of risk detection and analysis. A great deal of work has been done in this area. Solutions have been attempted using a wide variety of methods (e.g., optical flow, Kalman filtering, hidden Markov models, etc.) and modalities (e.g., single camera, stereo, infra-red, etc.). In addition, there has been work in multiple aspects of the issue, including single pedestrian tracking, group tracking, and detecting dropped objects. For surveillance applications, tracking is the fundamental component. Kalman filters have been used extensively for tracking in many domains. It is worthy of note that most applications used only a linear Kalman filter approach. It seems that this was sufficient for many problems. Many applications could model 2D or near 2D motion exclusively.

II. DESCRIPTION OF WORK

This research developed in two parts: Human detection and tracking based on position and velocity. Specifically, we built upon a framework of code and this code tracked objects appearing in a digitized video sequence with the use of a mixture of Gaussians for background/foreground segmentation and a Kalman filter for tracking. All experiments were run on 320x240 pixel resolution images on a computer with a Pentium IV 2 GHz single processor and 1 GB of RAM. The goal of this stage was to segment and extract the image of each pedestrian from all appearances in the image sequence. This "pedestrian image sequence" data could then be used in the later stages of the system to provide information to the motion recognition components to classify the pedestrian motion we developed routines to accomplish three things:

1. Establish a stable oversized bounding box around pedestrians tracked smoothly throughout video sequence.
2. Grab the image of the pedestrian within the bounding box and save it.
3. Combine the individual images into movie files.

This module could then track a pedestrian and generate single image snapshots or movies of the pedestrian's motion. Figure 4 below shows some example image sequences generated

III. HUMAN ACTIVITY RECOGNITION BASED ON PEDESTRIAN POSITION AND VELOCITY

This component estimates the pedestrian motion based on the speed and position of the pedestrian. The basic assumption is that much of the pedestrian's activities can be ascertained by measuring these simple aspects. Measuring these values provides several advantages over articulated motion analysis: these measurements can be made in real time and are far more robust to noise and poor image quality. In addition, for our purposes, if a pedestrian is moving in an area that is off limits, that should be flagged as a warning. In this circumstance, the type of motion is generally irrelevant. This process had several components:

1. Track each pedestrian throughout scene using the Kalman filter estimates.
2. Record the position and velocity state.
3. Develop a position and velocity path characteristic for pedestrian. This was done using the Kalman filter prediction of future state.

A. Detection of Moving Object

To achieve object detection, there are two methods or models which are widely used by researchers. One of them is background subtraction model. This model basically subtracts image between the reference image and current image. The reference image usually comes from the background image which only consists of the background. For current image, it consists of the same background as the reference with an addition of moving people within it the background subtraction model was based on the equation 3.1 below. Assume N as the least intensity value and M as the maximum intensity value for pixel *x*, N and M values will be subtracted from the image I to obtain the largest inter-frame absolute difference D.

$M(x) - I(x) = D(x)$ or

$$N(x) - I(x) = D(x) \qquad (3.1)$$

The difference (D) of pixel *x* will be the detected foreground or detected moving object. The moving object (foreground) detection is stated in equation 3.2 where pixel *x* from image I is a foreground if D is larger than the value being subtracted from N or M.

$M(x) - I(x) > D(x)$ or

$$N(x) - I(x) > D(x) \qquad (3.2)$$

In this system, the foreground detection or object detection is done by giving the maximum M, minimum N and median of the largest inter-frame absolute different dµ images over the entire image that was represented by the background scene model B. The foreground I(x) from the image of pixel *x* if:

$$B(x) = \left\{ \begin{array}{l} 0(background \quad) \left\{ \begin{array}{l} I(x) - M(x) < kd_{\mu} \\ vI(x) - N(x) < kd_{\mu} \end{array} \right\} \\ 1(\ foreground \quad) - - otherwise \end{array} \right\} (3.3)$$

B. Background Subtraction Process

Background subtraction model as given in equation 3.1 and 3.2 was used for the next step to obtain all moving objects present in the image frame. This process involves taking the difference between the reference image frame and current image frame. Both image frames are in grayscale color. In MATLAB operation this involves pixel by pixels processing. The output after the background subtraction process will only contain the moving object without any background as shown in Figure 1 below.

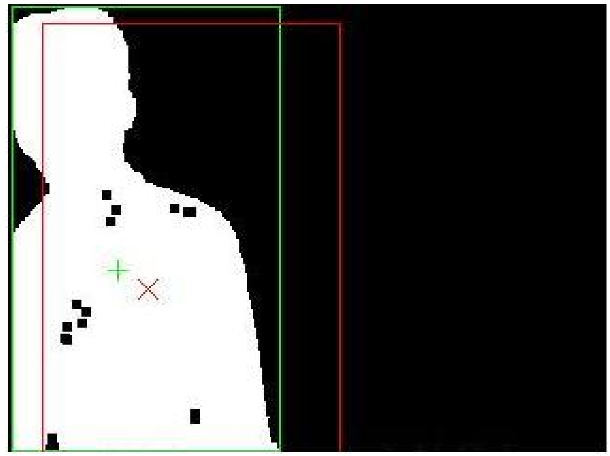


Figure 1: Object detection after background subtraction process

IV. LINEAR MODEL AND KALMAN FILTER

This filter based on linear dynamical systems discredited in time domain. They are modeled on Morkov chain built on a linear operator perturbed by Gaussian noise. The state of the system is represented as a vector of real numbers. At each discrete time increment, a linear operator is applied to the state to generate a new state, with some noise mixed in, and some information from contents of systems. They are recursive filters, means that only the estimated state from previous time state and current measurement are needed to complete the estimate for current state. It has two distinct phases, a prediction and updating. The measurement information for current time step is used to refine this prediction to arrive at new more accurate estimate. Kalman filter alone performs well when the target dynamics is linear and matches the model used in the filter.

State equation:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \qquad (4.1)$$

Output equation:

$$y_k = cx_k + z_k \qquad (4.2)$$

In the above equations A, B, and C are matrices; k is the time index; x is called the state of the system u is a known input to the system; y is the measured output; and w and z are the noise. The variable w is called the Process noise and z is called the measurement noise. Each of these quantities are (in general) vectors and therefore contain more than one element. The vector x contains all of the information about the present state of the System, but x cannot be directly measured instead measure y, which is a function of x that is corrupted by the noise z. y can be use to help us obtain an estimate of x, but we cannot necessarily take the information from y at face value because it is corrupted by noise. the information which is presents to a certain extent, but cannot afford to grant it as perfectly accurate. For example, suppose a model of vehicle going in a straight line. The state consists of the vehicle position p and velocity v. The input u is the commanded acceleration and the output y is the measured position. The acceleration can be changed and measured the position at every T seconds. In this case, elementary laws of physics say that the velocity v will be governed by the following equation

$$v_{k+1} = v_k + Tu_k \qquad (4.3)$$

That is, the velocity one time-step from now (T seconds from now) will be equal to the present velocity plus the commanded acceleration multiplied by T. But the previous equation does not give a precise value for v_{k+1}. Instead, the velocity will be perturbed by noise due to gusts of wind, potholes, and other unfortunate realities. The velocity noise is a random variable that changes with time. So a more realistic equation for v would be:

$V_{k+1} = V_k + Tuk + V_k^{\sim}$ (4.4)

where $Vk^{\sim}$ is the velocity noise. A similar equation can be derived for the position p:

$P_{k+1} = P_k + TV_k + 1/2(T^2 u_k) + P_k)$ (4.5)

where $P_k^{\sim}$ is the position noise. Now we can define a state vector x that consists of position and velocity: Finally, knowing that the measured output is equal to the position, we can write our linear system equations as follows:

$y_k = C\,x_k + z_k$ (4.6)

z_k is the measurement noise due to such things as instrumentation errors. If the vehicle needs to be controlled with some sort of feedback system, an accurate estimate of the position p and the velocity v is needed. In other words, a way to estimate the state x is needed. This is where the Kalman filter comes in.

Prediction:

$x_k^{\wedge} = Ax^{\wedge}_{k-1} + w_{k-1}$ (4.7)

$P_k = AP_{k-1}A^T + Q - AP_{k-1}S^{-1}_{k-1}P_{k-1}A^T$ (4.8)

Update

$K_k = AP_k (P_k + R)^{-1}$ (4.9)

$X^{\wedge}_{k+1} = Ax_k^{\wedge} + K_k(y_{k+1} - Ax_k^{\wedge})$ (4.10)

$P_{k+1} = (I - K_k S_k P_{k)})$ (4.11)

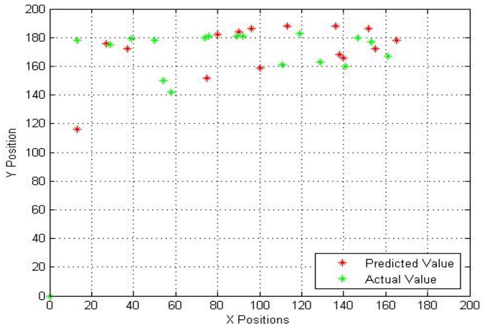
That's the Kalman filter. It consists of equations, each involving matrix manipulation. In the above equations, a –1 superscript indicates matrix inversion and a T superscript indicates matrix transposition. The K matrix is called the Kalman gain, and the P matrix is called the estimation error covariance. The state estimate 4.9 equation is fairly intuitive. The first term used to derive the state estimate at time k + 1 is just A times the state estimate at time k, plus B times the known input at time k. This would be the state estimate if a measurement is absent . In other words, the state estimate would propagate in time just like the state vector in the system model. The second term in the equation is called the correction term and it represents the amount by which to correct the propagated state estimate due to our measurement. Inspection of the K equation shows that if the measurement noise is large, Sz will be large, so K will be small and credibility to the measurement y is not given when computing the next. On the other hand, if the measurement noise is small, Sz will be small, so K will be large and a lot of credibility to the measurement is given when computing the next.

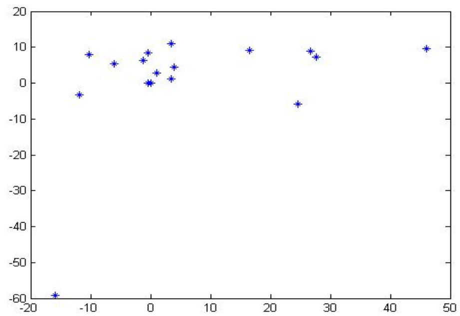
V. EXPERIMENTAL RESULTS

Table 1 shows the actual and predicted values of positions of object

Actual values		Predicted values		Residue	
49.6957	178.1292	49.5401	178.1400	0.1556	-0.0108
73.8278	180.4705	74.2706	180.4533	-0.4428	0.0172
79.8195	182.0264	76.3483	180.8123	3.4712	1.2141
90.3713	183.7930	89.3577	180.9525	1.0136	2.8405
95.8887	185.7249	91.9146	181.1766	3.9741	4.5483
113.2230	188.3991	119.1986	182.9622	-5.9756	5.4369
136.4907	187.5993	146.6808	179.5698	-10.1901	8.0295
152.4753	185.6527	152.9317	177.2071	-0.4564	8.4456
164.9695	177.7554	161.4480	166.6516	3.5215	11.1038
155.3507	171.5121	128.5869	162.5106	26.7638	9.0015
138.2796	167.7668	110.6707	160.5341	27.6089	7.2327
140.2382	165.8109	141.3830	159.5543	-1.1448	6.2566
99.9278	159.1645	53.9195	149.5631	46.0083	9.6014
74.5232	151.5484	58.0411	142.4037	16.4821	9.1447
37.2700	172.1738	12.6425	177.9546	24.6275	-5.7808
13.1530	115.8104	28.9823	174.9779	-15.8293	-59.1675
27.2456	175.6311	39.0311	178.9119	-11.7855	-3.2808

Graph 1 showing predicted & Actual Values. Graph 2 Difference between actual & predicted values





We tested the system in an outdoor environment, when a human being comes in a camera vision our system detect and track the same. The figures 2, 3 and 4 below shown are the surveillance images taken with the camera. We have detected and tracked the person by background subtraction technique and by the Kalman filter. The green circle shows the actual location and red circle shows predicted location of human being by Kalman filter. Statistically we have shown the actual and predicted location in Graph 1 and difference between actual and predicted shown in graph 2.

Fig.2

Fig.3

Fig.4

VI. CONCLUSIONS

This paper proposes a method for tracking objects in image sequences using background subtraction technique and Kalman filter respectively. To adapt the template to changes of object appearance during tracking, multi-value appearance features at pixels are smoothed temporally by robust and adaptive Kalman filters, allowing for the accurate detection of the object. The residual information is exploited to tune the filter parameters automatically.

Kalman predictors can be used for motion detection of moving objects. The Kalman model is adaptive model which acquires the accuracy from previous steps. Results are showing that Kalman predicted values and actual values are very near to each other.

REFERENCES

[1] J.P.b1acCormick and A. Blake, "A probabilistic exclusive principle for tracking multiple objects," in ICCVYg, 1999, pp. 572-578.

[2] H. Tao, H.S. Sawliney, and R. Kninar, "A sampling algorithm for tracking multiple objects," in Vision Algorithms 99, 1999

[3] C.Hue, J.P.LeCadre, andP.Perez,"Multiple object tracking by particle filtering," IEEE Transc. On Aerospace and Electronics Systems, vol. 38, no. 3: pp. 791-812. .July 2000

[4] D.B. Reid, "An algorithm for tracking multiple targets," AC, vol. 24, no. 6, pp. 843-854, December 1979.

[5] C. Stauffer and W.E.L. Grinwon, "Learning patterns of activity using real-time tracking," PAMI, vol. 22, no. 8, pp. 747-757, August 2000.

[6] MathiesL, Szeliski, RandkanadeT.1989.kalman filter based algorithms for estimating depth. from image sequences international journal computer vision 3,3 209-238.

[7] Ali A, and Agrawal J.'segmentation and recognition of continuous human activity .In IEEE workshop on detection and recogn of events in video 28-35.

[8] Agrawal J and CAI Q 1999'Human motion analysis's review computer vision image Understand 73, 3,428-440.

[9] Mei Hail Amit sellit wei Hua, A Detection based object tracking method 2004

[10] Neeti A. Ogale, A survey of techniques for human detection from video.

[11] Liang Wang, Weiming Hu, Tieniu TanRecent Developments in Human Motion Analysis National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080

[12] Sonka, Hlavac & Boyle, Digital image Processing and Computer Vision, Cengage Learning publication, 2008