Improving Object Tracking Accuracy in Video Sequences Subject to Noise and Occlusion Impediments by Combining Feature Tracking with Kalman Filtering

Mark Heimbach
Department of Electrical Engineering
Santa Clara University
Santa Clara, California 95053
Email: mheimbach@scu.edu

Kamak Ebadi
Department of Electrical Engineering
Santa Clara University
Santa Clara, California 95053
Email: kebadi@scu.edu

Sally Wood
Department of Electrical Engineering
Santa Clara University
Santa Clara, California 95053
Email: swood@scu.edu

***Abstract*—Automated object tracking in video sequences has been applied in areas such as security and surveillance, traffic control, medical image processing and video communications. To maintain robust precision in matching our algorithm's position estimates with ground truth, feature detection accuracy and confidence measures are needed by tracking algorithms. Experiments are conducted to quantify how feature descriptors used for tracking are degraded. A Kalman filter is then used to enhance accuracy. In our application, Kalman covariance parameters are continually tuned using a confidence level obtained based on object descriptor robustness. Because the Kalman algorithm converges quickly and does not require prior training, it is ideally suited for real-time object tracking.**

***Index Terms*—Kalman, Histogram of Oriented Gradients, Occlusion, Covariance, Object Tracking.**

I. Introduction

Feature detection based object tracking can be difficult due to occlusions and noise impediments such as Gaussian noise, blurring, saturation, salt-and-pepper noise and illumination variability. In our implementation, the Histogram of Oriented Gradients (HOG) [1] is used to identify object features in a video frame. However, there are certain vulnerabilities for HOG based detectors. These include loss of track due to noise impediments, occlusions and false detections.

Using HOG feature detection, a ground truth based gradient vector is used for correlation purposes. Ambiguity can occur when the maximum HOG response is close in magnitude to other responses located at different positions within the image. The presence of noise exacerbates this problem, increasing possibilities for false object detections and loss of track.

In real-time object tracking, only a subsection of the video frame is searched. This reduces computational complexity, making real-time tracking possible. Other approaches have been successful, but at the expense of computational complexity [2], [3]. If the tracked object is lost, it cannot be easily recovered.

II. Kalman Filter

The Kalman Filter [4] is used for predicting a linear system in the presence of Gaussian distributed noise. The Kalman filter process and measurement equations (1-2) use the respective update equations (3-7). For our tracking application, the state vector in (8) includes position and velocity.

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

$$z_k = Hx_k + v_k \tag{2}$$

Time Update Equations

$$\widehat{x}_k = A\widehat{x}_{k-1} \tag{3}$$

$$P_k^- = A\widehat{P}_{k-1}A^T + Q \tag{4}$$

Measurement Update Equations

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{5}$$

$$\widehat{x}_k = \widehat{x}_k^- + K_k(z_k - H\widehat{x}_k^-) \tag{6}$$

$$P_k = (I - K_k H)P_k^- \tag{7}$$

with the following parameters defined below-

$$X_k = \begin{vmatrix} X & Y & \dot{X} & \dot{Y} \end{vmatrix}^T \tag{8}$$

$$A = \begin{vmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & dt & 0 \\ 0 & 0 & 0 & dt \end{vmatrix} \quad H = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix} \tag{9}$$

$$Q = E\left[\omega_k \omega_k^T\right] = \sigma_w^2 \frac{dt}{4} \begin{vmatrix} 1 & 0 & 6 & 0 \\ 0 & 1 & 0 & 6 \\ 6 & 0 & 2 & 0 \\ 0 & 6 & 0 & 2 \end{vmatrix} \tag{10}$$

$$R = E\left[\nu_k \nu_k^T\right] = \sigma_\nu^2 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \tag{11}$$

$$P_k = E\left[(x_k - \widehat{x}_k)(x_k - \widehat{x}_k)^T\right] \tag{12}$$

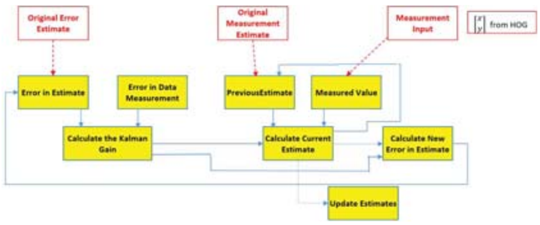

Fig. 1. Diagram depicting steps used by Kalman filter.

The Kalman filter is a linear estimator which minimizes mean square error between estimate and actual position. Real-time estimates for system parameters such as position and velocity are made for each time increment. Each estimate is continually updated using measurements subject to noise impediments. A HOG feature detector provides the measurement for the tracked object position (see Fig. 1).

To illustrate Kalman filter tracking capabilities, several one dimensional plots were constructed showing tracking capabilities using an input signal with an abrupt position change. Fig. 2&3 are shown with a Gaussian noise added to the signal with standard deviation σ = 0.2 and 0.4 respectively.

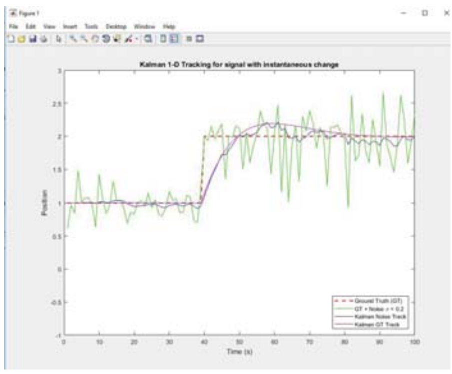

Fig. 2. Kalman filter tracking abrupt change (σ = 0.2)

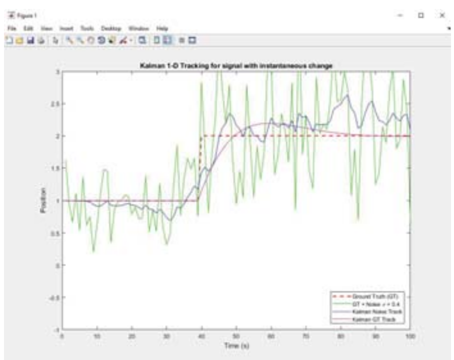

Fig. 3. Kalman filter tracking abrupt change (σ = 0.4)

One further observation is shown here whereby the Kalman measurement uncertainty ω_k is varied. Measurement uncertainty can be updated for each measurement based on the performance of our feature tracker. This will be discussed below.

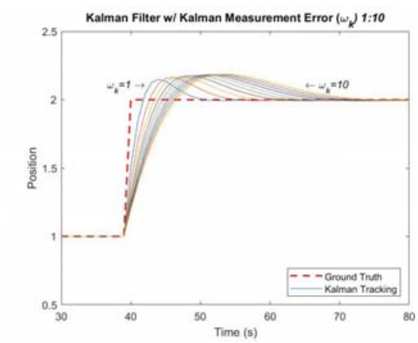

Fig. 4. Kalman tracker response vs changes in measurement noise uncertainty.

A. Histogram Oriented Gradients Feature Detector

Feature extraction is a fundamental step in object detection. The histogram oriented gradient method proposed by Dalal et al Dalal [5] has been widely accepted due to its relative insensitivity to rotation, deformation and illumination. The fundamental concept behind HOG is that an object shape in an image can be recognized by its edges. The HOG descriptor captures and encodes the object shape as the distribution of local intensity gradients or edge directions. A brief description of the process used in the HOG algorithm is shown below.

Gradients are calculated for each pixel in the image in both the x and y directions.

Gradient Vector Calculations

$$G_h(x,y) = f(x+1,y) - f(x-1,y) \ \ \forall x,y \tag{13}$$

$$G_v(x,y) = f(x,y+1) - f(x,y-1) \ \ \forall x,y \tag{14}$$

Gradient Magnitude

$$M(x,y) = \sqrt{G_h(x,y)^2 + G_v(x,y)^2)} \tag{15}$$

Gradient Direction

$$\theta(x,y) = tan^{-1}\left(\frac{G_v(x,y)}{G_h(x,y)}\right) \tag{16}$$

To minimize computation time for real-time tracking, only a subset of the image (detection window) is searched to determine object position. The detection window is continuously updated such that the object is centered inside the window in each video frame.

Each pixel of the image is filtered using the template [-1 0 1]. The image is organized into cells, each of which comprises 8x8 pixels. Each cell consists of 9 histogram bins. Four cells compose a block. Blocks are scanned across the image in both

Fig. 5. Image shows Detection Window "Patch (red box)". Target object shown in (blue box). HOG scan conducted in Detection Window only to facilitate real-time tracking.

the X and Y directions, allowing for a 50% overlap. The image with corresponding detection window and target area are shown in the figure above. The histogram binning process and HOG structure is illustrated in the figure below.

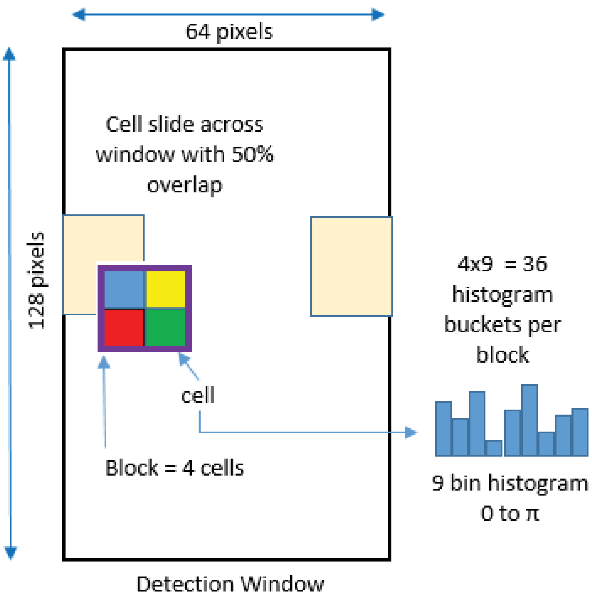

Fig. 6. HOG calculation algorithm

Once the HOG detection process completes its scan across the detection window, a response is given for each cell area. The maximum response shows object position. The value of the maximum response varies according to the degree of correlation. It is this information that is used to enhance Kalman filter accuracy. Attempts have been made to use Mean Shift descriptors with Kalman filters [6]. In this case Kalman was used to center the target window only. Other descriptors have been proposed to use with Kalman filtering. Invariant feature descriptors (IFD) [7] have been used to mitigate occlusions. However, this methods is strictly heuristic in its approach [8].

B. Scaled Kalman Filter

An online database is used to experimentally verify the use of HOG response values to modify Kalman filter noise parameters Wu[9]. Fifty videos are made available, each with ground truth positions and noise impediments such as occlusion, blur and illumination variations, among others.

Using the 'jogging' video contained in the dataset, a plot was made showing the variations in ground truth positions as a function of each frame. This is shown in the figure below.

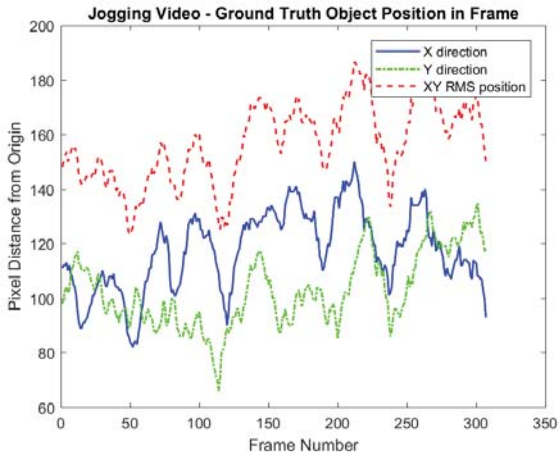

Fig. 7. Ground Truth position shown for each frame in the 'jogging' video.

The maximum HOG response is shown for each frame in the figure below. The maximum HOG response and the position error are inversely proportional.

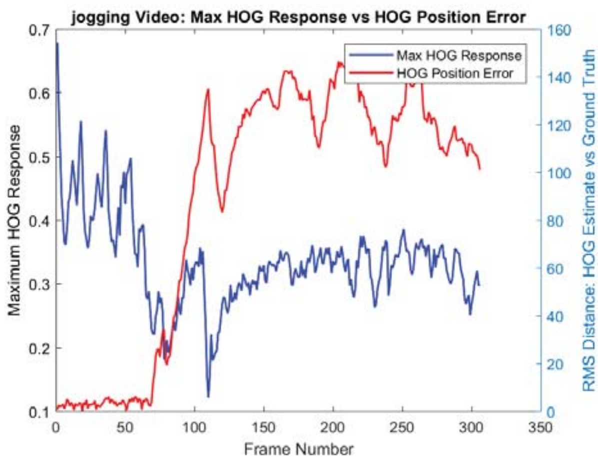

Fig. 8. Maximum HOG response and position errors shown for each frame. An occlusion occurs during the frames when the HOG response decreases, at which point track is lost.

C. Simulation Results

Image sequences are shown where the tracked object is occluded in the ”jogging” video. In this sequence, the jogger is temporarily occluded by an obstruction. The first set of images show track loss when a HOG-only feature detector is used. The second set of images depict a successful track using Kalman filtering to supplement the HOG detector in tracking the object of interest.

Bounding boxes are superimposed over each video that reflect the ground truth object position, HOG estimates and Kalman filter estimates. The target window is also shown here. Once the object moves beyond the target window, tracking is lost and cannot be recovered without reinitialization.

Color codes for each bounding box are as follows:

- Detection Window - Purple
- Ground Truth - Red
- HOG Estimate - Green
- Kalman Tracker - Yellow *(when used)*

Fig. 9. In jogging video, a Kalman filter is not used. Ground Truth positions (red) and HOG estimates (green) are each represented by a bounding box. Track is lost due to occlusion. Once the object leaves the detection window (purple), track cannot be easily reestablished.

Fig. 10. In this frame sequence, a Kalman filter (yellow bounding box) is used. Both HOG estimates and Ground Truth positions are also shown with bounding boxes. Track is maintained during occlusion event.

The video sequences show the Kalman filter (with emphasis on minimal process noise) tracks the Ground Truth position more accurately during and after the occlusion event. The HOG maximum response in Fig 8 shows an abrupt decrease in correlation accuracy (see frames 80-110) that indicates a reduced confidence level in the measurement. This information is used with the Kalman filter to maintain track.

III. Conclusion

The histogram oriented gradients feature detection operates by correlating a reference (ground truth) vector with gradient vectors measured in subsequent frames. The maximum HOG response (or position where the maximum vector correlation is measured) becomes the new object position estimate. In our approach, we have used the degree of correlation of that maximum HOG response as a measure of the confidence level that the estimate is accurate. The measurement confidence is used to scale the Kalman measurement noise variance to enhance its tracking capabilities. Here we have shown how the confidence we have in our HOG descriptor can be used to improve measurement robustness for occlusion events.

References

[1] Gao Ming Xu Fen. Human detection and tracking based on hog and particle filter. In *2010 3rd International Congress on Image and Signal Processing (CISP2010)*, pages 4293–4302, 2010.

[2] Han B. Nam H. Learning multi-domain convolutional neural networks for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4293–4302, 2016.

[3] Smeulders A.W. Tao R., Gavves E. Siamese instance search for tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1420–1429, 2016.

[4] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[6] Men Guo-zun Zhao Jie, Qiao Wen. An approach based on mean shift and kalman filter for target tracking under occlusion. In *Proceedings of the Eithth International Conference on Machine Learning and Cybernetics*, 2009.

[7] Koo V.C. Ong L.Y., Lau S.H. A new approach of local feature descriptors using moment invariants. In *Journal of COmputer Science*, pages 2538–2547, 2014.

[8] Koo Voon-Chet Ong Lee-Yeng, Lau Siong-Hoe. Performance of invariant feature descriptors with adaptive prediction in occlusion handling. In *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*, pages 385–388, 2017.

[9] Yang Ming-Hsuan Wu Yi, Lim Jongwoo. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418, 2013.