# Resolving Occlusion Ambiguity by Combining Kalman Tracking with Feature Tracking for Image Sequences

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*—An improved method for mitigating occlusion in object tracking is proposed in this paper. Using more traditional methods of object tracking and feature detection, a novel scheme is developed based on the use of multiple tracking methods which operate in parallel to minimize the effect of occlusion. A popular feature detection algorithm called Histogram Oriented Gradients (HOG) is used as our baseline tracker. Its ability to detect and track objects during occlusion is then enhanced with a Kalman filter. State variables for Kalman include position and velocity of the object. Additionally, we introduce a new term called a Correlation Constant C(k) which makes use of a HOG trackers noise distribution to minimize the process and measurement variance of the Kalman filter. An online video database is used to experimentally verify our proposed algorithm [1]. Each video frame is provided with ground truth coordinates for the object being tracked. Results were developed in Matlab using online code developed by Henriques [2] . Experimental results show that our proposed algorithm is effective in solving the occlusion problem.**

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

## I. Introduction

Object tracking is important for various applications such as video surveillance, vehicle navigation and robotics. However, object tracking can be difficult due to challenges such as illumination variability, occlusions, outliers, object scaling and camera motion. Here we focus on mitigating occlusions as impediments to tracking by looking at combining tracking methods. These methods then operate in parallel to trackers based on feature detection. Kalman filtering is examined as one possible tracker that can predict object motion in the presence of noise. This method is applicable to many physical scenarios where motion in one direction tends to proceed in the same direction. For example, a car moving in one direction will not suddenly proceed to travel in the opposite direction. Accurate tracking depends of the trackers ability to maintain a close relationship to the object of interest. When track is lost due to occlusion, the appearance of outliers or other anomalies, the track can be lost. In many real-time tracking applications, a search window or patch is contained within the frame to minimize the search area. When algorithms are used to search for the object only in these areas, the track can not only be lost, but can also be unrecoverable. To model the improvements to tracking with the Kalman filter, some standard tracking methods will be used. This includes the use of the Histogram of Oriented Gradients (HOG) as a method for feature detection.

## II. Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) feature detector is used widely in characterizing objects by examining the distribution of intensity gradients or edge directions. HOG was introduced by Dalal [3] at the 2005 CVPR conference. The detector is trained using object images where a detection window performs a scan, identifying pixel-to-pixel value changes in both the vertical and horizontal directions. These pixel changes are then computed according to magnitude and direction of change. The changes are then categorized into a histogram with each bin organized by directional orientation. Finally, a column vector is calculated containing all the histogram values accumulated during the image scan. These vectors are then cross-correlated with actual images to identify and track the object of interest.

Training is initially conducted within the first video frame to identify important features of the object to be tracked. Correlation filtering is then used to measure the degree to which a set of features obtained by training corresponds to the object as it is being tracked. The trained filter is swept across each patch window. The location where the corresponding correlation is highest is designated to be the position of the object. The Histogram of Oriented Gradients method is used as a feature detector.

The first step in our process is to establish a reference, which is done in the first frame. In this instance, we are tracking a specific basketball player. The detection window patch and object target are shown.

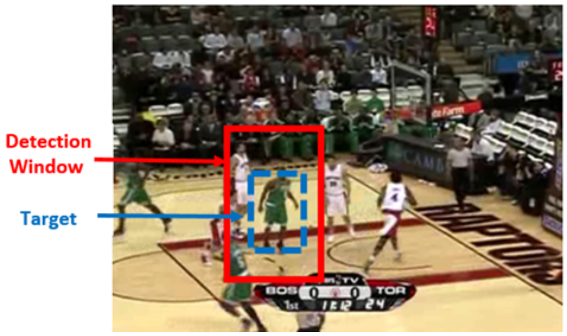


Fig. 1. Reference Frame: Image shows Detection Window "Patch" (red box) and Target Object (blue box).

The process for computing HoG is shown below. Gradient filters are applied across each pixel to determine edges and other features used in the tracking process.

*Gradient Vector Calculations*

$$G_h(x,y) = f(x+1,y) - f(x-1,y) \;\; \forall x,y$$
$$G_v(x,y) = f(x,y+1) - f(x,y-1) \;\; \forall x,y$$

*Gradient Magnitude*

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

*Gradient Direction*

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

### A. HoG Response

Once the HoG gradient filter is developed in the reference frame, the HOG filter is applied to the patch in all subsequent frames. Correlation between the reference HOG and subsequent frames is done for each cell in the patch. Here we define each cell to contain an 8 pixel by 8 pixel square. Two cells are grouped together to form a block used for normalization calculations. Normalization serves to mitigate the effect of illumination changes on the frame. The result of each HoG correlation is a real number. Once the correlation is conducted for each cell, the highest value HoG response indicates the target position within the patch. Given various noise impediments in the system, the maximum HoG response varies with each frame. The figure below shows the maximum HOG response as a function of video frame number. The HOG responses are normalized to a value of one.

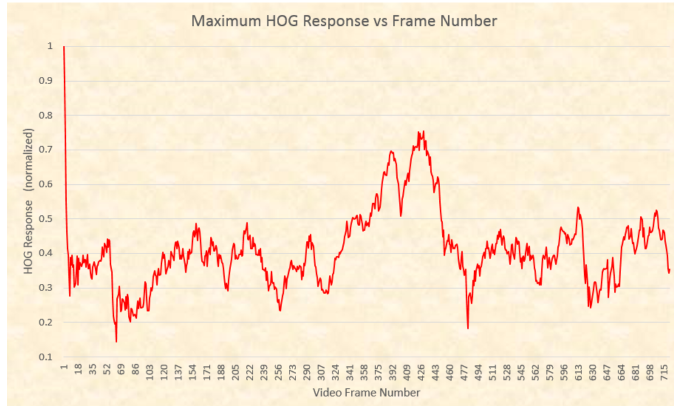


Fig. 2. Normalized HoG Maximum Response for each Frame in Video.

In can be seen from this plot that the value of the HOG reference correlation for the target in each frame is subject to variability. This information is used to develop a confidence factor proportional to the distance between each maximum HOG response and the maximum HOG response from our reference frame. The correlation between trainer (reference) and target in subsequent frames is indicative of the confidence level that the detector is indeed tracking the object at the correct location in the frame. The confidence factor from the HOG tracker is used to improve the noise covariance estimate of the Kalman filter, additionally making the tracker more robust to occlusion events. The figure below shows the impact of occlusion on a HoG tracker without the Kalman filter.

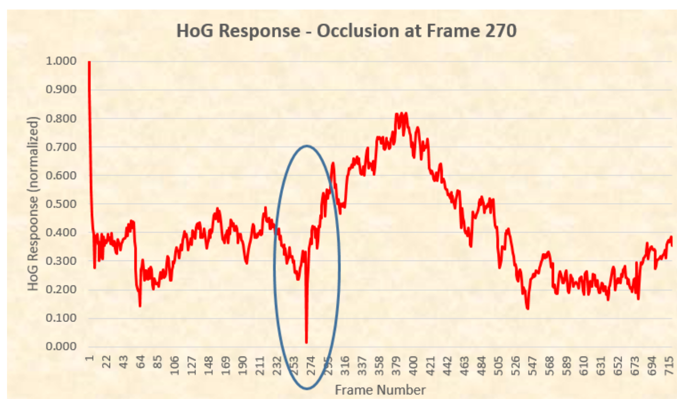


Fig. 3. Maximum HoG Response vs Frame Number. Occlusion occurs at Frame 270 (*highlighted by blue oval*)

HOG video frames along with their corresponding surface and contour plots are shown below. Shown in the first row are the reference frame, a typical frame (second row) and a video frame (third row) which is experiencing an occlusion event.

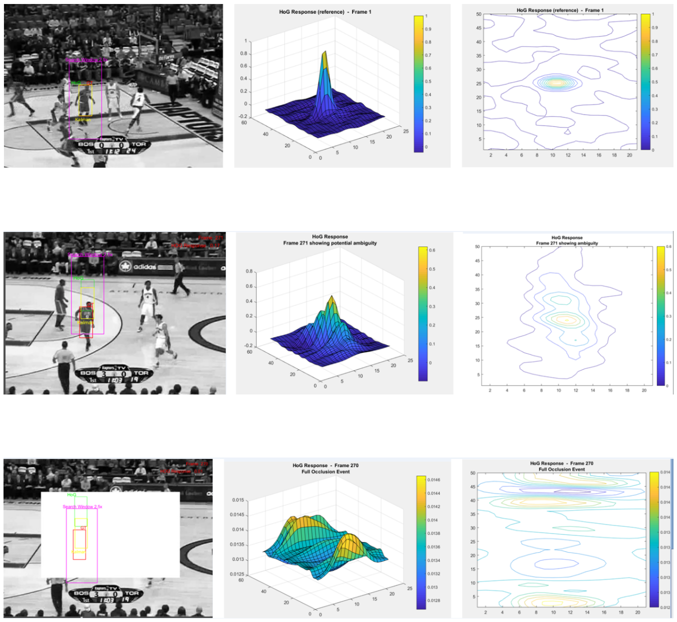

Fig. 4. Reference, Typical and Occlusion Frames shown with corresponding Surface and Contour Plots respectively. A higher Surface plot peak is indicative of a higher confidence level in tracker positioning.

## III. Kalman Filter

The Kalman Equations used here are listed below. the First equation in the Kalman Model indicates the updated state equation for the next object position based on the object's prior state, prior measurement and prediction noise. The second equation adds measurement noise to the object state estimate, predicting the new position.

**Kalman Model**

$$X_t = A * X_{t-1} + B * u$$

$$Y_t = C * X_t + E_z$$

**State Variables**

$$\left[ X \; Y \; \dot{X} \; \dot{Y} \right]^T$$

$$A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} \frac{dt^2}{2} \\ \frac{dt^2}{2} \\ dt \\ dt \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

**Kalman Equations**

$$\widehat{X}_{t|t-1} = A * \widehat{X}_{t-1|t-1} + B * u_{t-1}$$

$$P_{t|t-1} = A * P_{t-1|t-1} * A^T + E_x$$

$$K_t = P_{t|t-1} * C^T * (C * P * C^T + E_z)^{-1}$$

$$\widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t * (Z_t - C * \widehat{X}_{t|t-1})$$

$$P_{t|t} = (I - K_t * C) * P_{t|t-1}$$

$$E_x = \begin{bmatrix} \frac{dt^4}{4} & 0 & \frac{dt^3}{2} & 0 \\ 0 & \frac{dt^4}{4} & 0 & \frac{dt^3}{2} \\ \frac{dt^3}{2} & 0 & dt^2 & 0 \\ 0 & \frac{dt^3}{2} & 0 & dt^2 \end{bmatrix} \quad E_z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

*where $Z_t$ is HOG/KCF measurement at time t*

Fig. 5. Kalman Equations. Position and Velocity used as State Variables.

Noise is assumed to be Gaussian. $E_x$ is the covariance matrix for position and velocity while $E_z$ represents measurement noise.

### A. Kalman Filter using Correlation Constant

The Kalman filter is used to supplement the HoG tracker to mitigate problems such as occlusion and false detections due to outliers. We use the information provided by the confidence level of the HoG tracker to enhance Kalman tracker ability. A model for the use of the HoG tracker supporting the Kalman filter's ability to mitigate occlusion is shown here.

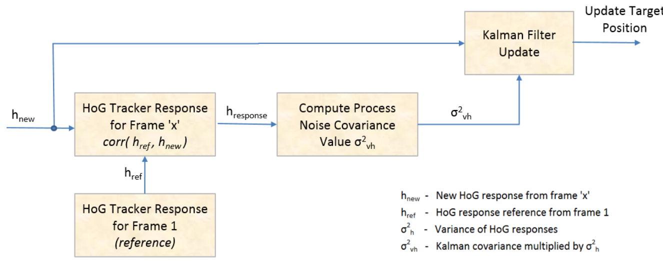


Fig. 6. Kalman Filter Model using HoG Confidence Information to Mitigate Occlusion.

By using the HoG information as a confidence factor, we are able to improve the Kalman Filter's noise profile to provide an improved tracking capability. By multiplying the variances of both the HoG response with the Kalman Filter's process noise variance, an improved overall noise variance can be obtained. See the figure below where two variances are multiplied together to form an improved mean and variance profile.

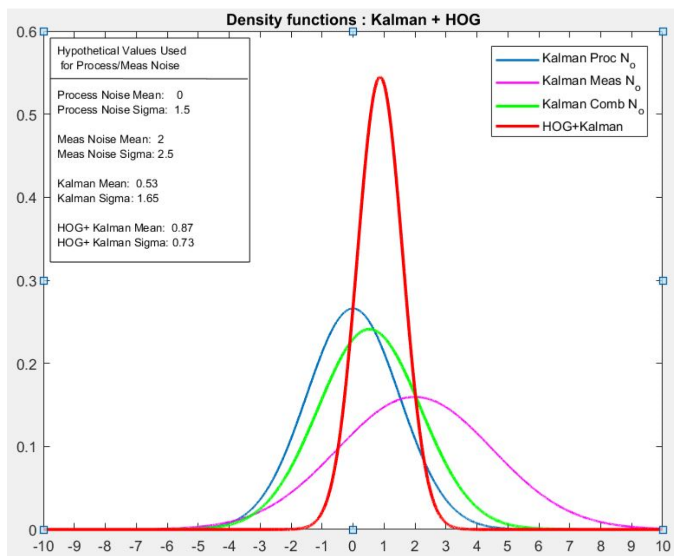


Fig. 7. Model depicting mean/variance improvements by multiplying each individual variance.

### B. Kalman Simulation

A Kalman filter is used in addition to the HoG feature-based tracker to mitigate problems such as occlusion, changes in illumination and false detections due to outliers. Such a method was applied successfully by Bogun, et al [4]. We extend the Bogun implementation by using HoG information to supplement the Kalman filter tracking estimations. Zhao, et al [5] uses Mean Shift and Kalman filtering for tracking. The method differs from our current approach in that only the search window is updated by the Kalman filter.

The Kalman filter is updated based on both process and measurement information for each frame. With a complete occlusion, the measurement distance of the HoG response between the reference and occluded frame is high. This indicates a low confidence level that the measurement is tracking correctly. In this situation, emphasis is given to the Kalman process information, discounting the measurement equation

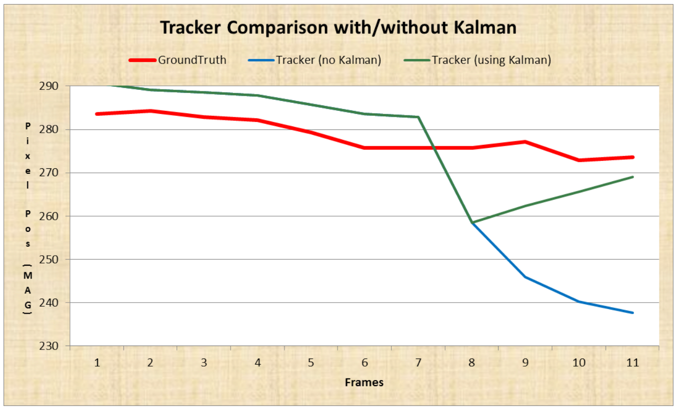


Fig. 8. Occlusion occurs in Frame 7 in this case. HoG tracker using the Kalman Filter provides a position estimation which is closer to Ground Truth.

The figure above shows how the Kalman filter (with emphasis on minimal process noise) stays closer to the Ground Truth position of the object during and after an occlusion event. This is compared to the HoG tracker working alone.

## IV. Conclusion

An occlusion example is shown below. Frames 269, 270 (with occlusion) and 271 are shown side by side. Bounding boxes are displayed and labelled appropriately (Search Window, HoG tracker, Ground Truth and Kalman Filter). It can be seen that the HoG tracker diverges substantially from the Ground Truth bounding box. The first set of figures show results when the Kalman filter is not used together with HoG. The second sequence uses both HoG and the Kalman filter. Observe the third frame (frame 271) in each sequence. The tracker is more closely aligned with Ground Truth when both HoG and Kalman trackers are used.

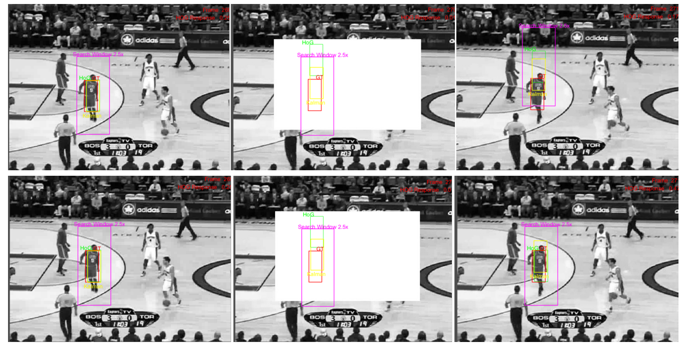

Fig. 9. Frames 269, 270 and 271. The Kalman filter is not used for the top sequence. After the occlusion, the Kalman tracker closely tracks the Ground Truth.

When an object of similar appearance to the target appears, the tracker using HoG only eventually loses track of the target. The tracker using HoG+Kalman maintains the correct track. See the pictures below for Frame 290.

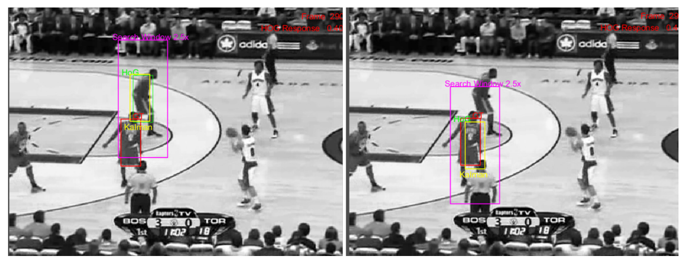

Fig. 10. Frame 290. Ten frames after the occlusion, the HoG-only tracker (left) begins to tracker the wrong target. HoG+Kalman tracker (right) follows the the correct target.


## Acknowledgment

The authors would like to thank Santa Clara University for its support of our participation in the Asilomar Conference.


## References


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

[2] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2015.

[3] 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.

[4] Ivan Bogun and Eraldo Ribeiro. Robstruck: Improving occlusion handling of structured tracking-by-detection using robust kalman filter. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 3479–3483. IEEE, 2016.

[5] Jie Zhao, Wen Qiao, and Guo-Zun Men. An approach based on mean shift and kalman filter for target tracking under occlusion. In *Machine Learning and Cybernetics, 2009 International Conference on*, volume 4, pages 2058–2062. IEEE, 2009.