Feature Prominence-Based Weighting Scheme for Video Tracking

Sema Candemir*
Dept. of Computer Science,
University of Missouri,
Columbia,MO,65211,USA.
candemirs@missouri.edu

Kannappan Palaniappan
Dept. of Computer Science,
University of Missouri,
Columbia,MO,65211,USA.
pal@missouri.edu

Filiz Bunyak
Dept. of Computer Science,
University of Missouri,
Columbia,MO,65211,USA.
bunyak@missouri.edu

Raghuveer M. Rao,
Army Research Lab.,
Adelphi, MD 20783, USA
raghuveer.rao@arl.army.mil

Guna Seetharaman,
Air Force Research Lab.,
Rome, NY 13441, USA
Gunasekaran.Seetharaman@rl.af.mil

ABSTRACT

This paper introduces a new mechanism called *Feature Prominence* to combine evidence from multiple feature operators for more reliable target detection and localization during video tracking. Feature prominence is measured using the statistical p-value estimated from a non-parametric local kernel density estimate of the a posteriori feature distribution. More prominent features have lower p-values and this ordering can be used to either discard low prominence features (high p-values) or reduce their weight during the feature fusion process to produce a more reliable fused feature likelihood map for locating the target at a subsequent time during tracking. The proposed feature fusion method is embedded into a test tracking system. Then, detection and tracking performance of the system is evaluated. Experimental results indicated that feature prominence outperforms several other feature fusion methods.

Keywords

Feature Fusion, Statistical Significance Measure, Object Tracking Wide Area Motion Imagery

1. INTRODUCTION

Automatic object tracking algorithms are a key component in the effective utilization of the large volume of aerial imagery that is routinely collected. Especially at the terabyte per hour data rates of wide area motion imagery, that is designed for large geospatial coverage. However, compared to standard full motion video analysis tracking in wide area large format video poses a number of challenges including larger object displacement due to low frame rate sampling, static and dynamic parallax occlusions, mosaicing errors, lower resolution targets, and less image contrast [12]. Figure 1 shows sample wide area imagery for Philadelphia and a vehicle trajectory.

Figure 1: Sample wide area motion imagery [12] with zoomed view showing a sample vehicle track as a yellow trajectory.

Several studies have shown that combining different features together produces a more robust tracker [15, 6]. In cluttered environments, feature fusion can make a powerful distinction between the background and the similar objects. Figure 2 shows two scenarios, first where color can help distinguish the object and in the second where shape can be a definitive feature.

Figure 2: Left: Targets can be differentiated and tracked by color (intensity) feature. Right: Intensity ambiguity may distract the tracker. Shape information helps the tracker to handle the intensity ambiguity.

*Corresponding author

ICVGIP '12, December 16-19, 2012, Mumbai, India

Although using multiple features improves the accuracy and the robustness of a tracker, it brings out an important challenge which is the relative importance of features during fusion. In the literature, there are several fusion methods including sum rule and product rule [9, 11]. One of the approaches of the sum rule is the naive fusion which fuses the feature information using equal weights. However, a given feature will perform differently depending on target and background appearance characteristics. Therefore, equally weighted fusion of features may decrease the tracking performance, if some features do not perform well in that environment. Another approach is the weighted sum fusion in which every feature is weighted dynamically according to the feature strength. The importance of each feature should be adapted to the changes in target pose and the surrounding background for tracking applications. Therefore, a dynamic weighting approach is better to fuse the feature information, in which high quality features receive higher weight compared to less reliable features in an adaptive manner. This adaptation improves the performance under changes that are not modeled by the tracker itself.

In this paper, we introduce *feature prominence* to combine evidence from multiple feature descriptors. Feature prominence is measured using the statistical p-value estimated from a non-parametric local kernel density estimate of the a posteriori feature distribution. More prominent features have lower p-values (more statistically significant features). A p-value (significance value based ordering) can be used to appropriately adjust weights during the feature fusion process to produce more reliable fused likelihood maps and for locating the target at a subsequent time during video tracking.

1.1 Test-Bed System

We embedded the proposed feature fusion method into a test-bed tracking system which fuses the features by a weighted sum (Figure 3). Then, we evaluate the detection and tracking performance of the system.

The target and local search region are modeled by using a set of feature descriptors. The features are intensity histogram , gradient magnitude histogram, shape index using hessian, normalized curvature index, histogram of hessian eigenvector orientations, gradient orientation histogram, local binary pattern histogram, intensity normalized cross correlation, gradient magnitude normalized cross correlation. System generates a likelihood map for each feature by comparing the target to the local search region. The likelihood maps are then fused with a weighted sum. The local maxima in the fused maps are determined as the target location. Target localization is accurate if and only if the fused maps can accurately pinpoint the target which makes fusion a critical step of the test-tracking system.

We compare the proposed system to two other adaptive feature weighting methods on the same tracking system. Experimental results indicated that feature prominence outperforms the other feature fusion methods in terms of target localization and target tracking. This paper is organized as follows: In Section 2, we review two dynamic feature fusion methods which have been used for target tracking. In Section 3, we introduce a *feature prominence*-based fusion method followed by experimental results and conclusions.

2. FUSION METHODS

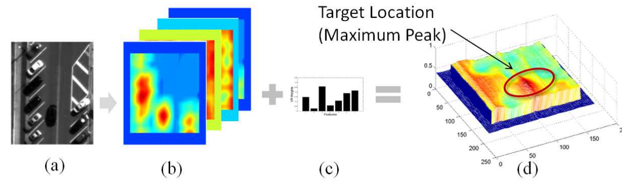

Figure 3: Feature probability maps are fused using a weighted sum and used for target localization: a) Search window, b) Feature likelihood maps, c) Feature weights, d) Fused likelihood map.

There are many potential approaches for robust feature fusion some of which are discussed in [8, 3, 10]. In this section, we review two dynamic fusion methods which adaptively determine the weights of likelihood maps during tracking.

2.1 Variance Ratio

Collins et al. [7] [16] state that the most discriminative feature between the target and the background are best for tracking the object. They adaptively weight the features according to the discrimination power between the target and the background using the two-class total to within class Variance Ratio (VR),

$$w_i \approx \frac{VR(L_i; fg, bg)}{\sum_i^n VR(L_i; fg, bg)} \tag{1}$$

where $VR(L_i; fg, bg)$ ratio of total variance to foreground and background within class variances,

$$VR(L_i, fg, bg) = \frac{var(L_i; (fg + bg)/2)}{var(L_i, fg) + var(L_i, bg)} \tag{2}$$

with fg representing the target distribution, and bg representing the background distribution in the likelihood map L_i for the i^{th} feature and is computed at the estimated target location in the previous frame.

Although, the Variance Ratio approach adaptively determines the weights of features, it is more suitable when all of the features are histogram based type. However, a tracking system may have other type of feature such as correlation based features which produces likelihood maps with sharper peaks in contrast to the likelihood maps produced by histogram-based features. The peak-like response underweights the correlation based features compared to histogram-based features as shown in Figure 4.

2.2 Distractor Index

Palaniappan et al. [13] proposed an improved method to fuse likelihood maps. They determine the number of local maxima within 90% of the maximum probability which gives the number of viable peaks for the i^{th} feature, $m_i \in [1, \infty)$. The Distractor Index based on the number of maxima is an inverse relationship given by $1 - (1/m_i)$. As the number of distractors increases (for a given feature) then its Distractor Index approaches one as shown in Figure 6. The Distractor Index based weighting scheme is given by $w_i \approx m_i^{-1} (\sum_{i=1}^n 1/m_i)^{-1}$. So high distractor index values will result in low weights for these unreliable features. An alternative Distractor Index measure that also monotonically increases with the number of peaks is $1 - e^{-\alpha(m_i - 1)^2}$

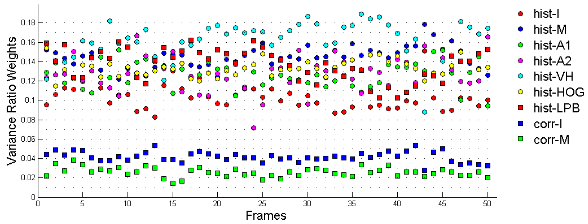

Figure 4: VR-based weights for different features over time shows that correlation features are consistently underweighted. The features are Intensity histogram (hist-I), Gradient magnitude histogram (hist-M), Shape index using Hessian (hist-A1), Normalized curvature index (hist-A2), Histogram of Hessian eigenvector orientations (hist-VH), Gradient orientation histogram (hist-HOG), Local binary Pattern histogram (hist-LBP), Intensity normalized cross correlation (corr-I), Gradient magnitude normalized cross correlation (corr-M). The reference frames of this graph are in Figure 5.

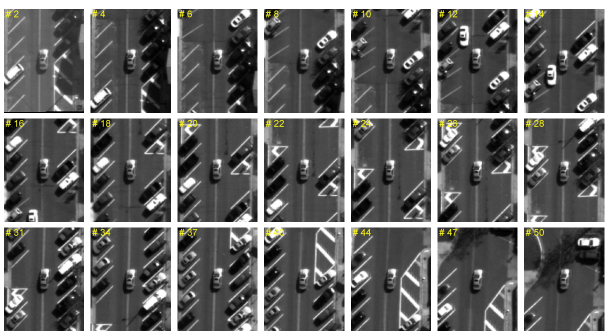

Figure 5: Reference frames of weight graphs in Figure 4, in Figure 7 and in Figure 9.

(Fig. 6).

Although experimentally the Distractor Index approach is usually more robust than the Variance Ratio method for fusing different types of features, it tends to over-weight the correlation type features. Figure **??** illustrates the weight relationship between histogram feature combination and correlation based features using the Distractor Index weighting scheme.

3. FUSION USING FEATURE PROMINENCE

In this paper, we propose *feature prominence* as a method for combining evidence from multiple feature operators for more reliable target detection and localization during video tracking. Feature prominence is measured using the statistical p-value estimated from a non-parametric local kernel density estimate of the a posteriori feature distribution. More prominent features have lower p-values and this ordering can be used to either discard low prominence features or reduce their weight during the feature fusion process to produce a more reliable fused feature likelihood map for locating the target at a subsequent time in video tracking.

In traditional statistics, statistical significance measures the randomness of an outcome in a hypothesis testing frame-

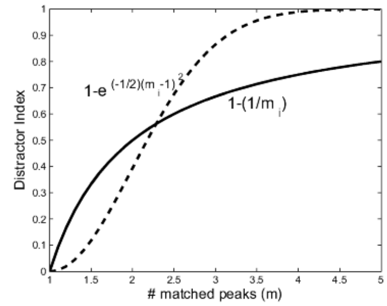

Figure 6: Distractor Index increases with greater number of peaks.

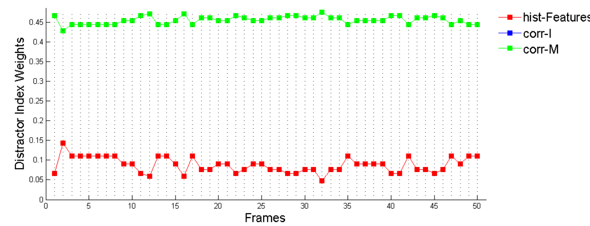

Figure 7: The weight relationship between histogram feature combination and correlation based features using distractor index weighting scheme. The reference frames of this graph are in Figure 5.

work. If the observed outcome of an experiment is statistically significant, this means that it is unlikely to have occurred by chance. Literature has some work in which statistical significance is used as a comparison measure between the outcomes of different distributions [4, 5]. In this paper, we use statistical significance as a measure of feature prominence.

At pixel location (x, y) let $f_i(x, y)$ be the value of feature i and $f_m(x, y)$ be the value of feature m. Let $p_i(f)$ be the probability distribution of feature i in the neighborhood of (x, y) and $P_i(f) = \int_{-\infty}^{f} p_i(x)dx$, be the associated cumulative distribution. Then if $P_i(f) < P_m(f)$ we say that feature i is more prominent than feature m and the weight associated with feature i should be higher. The test bed system models the target with a set of feature descriptors. Then it calculates a likelihood map for each feature by comparing the feature map of target with the feature map of local search window. Likelihood map has the same size with the local search window, and shows the probability of each pixel being part of target. The system fuses the likelihood maps with a weighted sum, following which locates the target as the local maxima of fused map. System calculates the likelihood maps for each frame and decides the target location for the next frame. Likelihood maps at each frame form the probability distribution of features. The probability of target location (x, y) is determined as the $f_i(x, y)$ value of feature i. The location of $f_i(x, y)$ among the other probability values is the significance value of feature i, which indicates the separability of associated feature for that particular frame. If $f_i(x, y)$ has a similar value to the distribution probabilities, then the cumulative distribution $P_i(f)$ will get higher value. The system assigns weight of feature i by formulating the weight as follows:

$$w_i \approx \frac{1 - P_i(f)}{\sum_i^n (1 - P_i(f))} \tag{3}$$

The probability distributions of features depend on the image and feature type. Figure 8 illustrates the weighting procedure using three probability distributions of uniform, gaussian and exponential. The red points on the distributions are feature values of the target (at a previous time step). When we evaluate each feature value using their cumulative distributions (Fig. 8.d), each value is assigned a significance value. The smallest significance value will be associated with the most prominent feature. We combine individual likelihood maps according to feature prominence weights using statistical significance values. More prominent features receive higher weight than less prominent features.

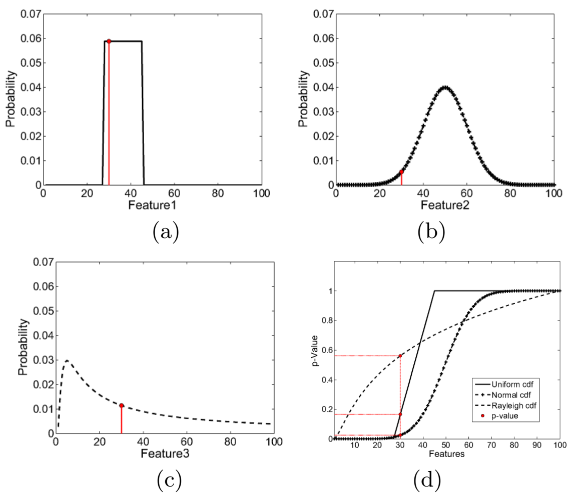

Figure 8: Three different feature probability distributions are shown: (a) Uniform for Feature 1, (b) Gaussian for Feature 2 and (c) Rayleigh for Feature 3. The red line marks the observed value of each feature in the sliding window. The cumulative distributions of the three features are plotted together in (d). The most prominent feature in this three feature example is Feature 2.

The feature weights with feature prominence weighting scheme are shown in Figure 9.

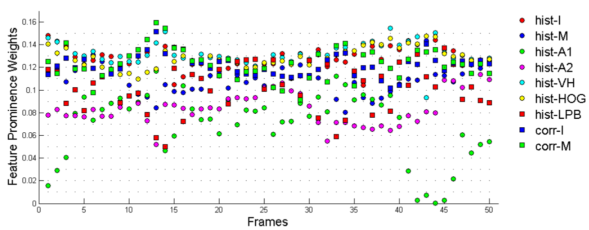

Figure 9: The weight relationship between histogram feature combination and correlation based features using feature prominence weighting scheme. The reference frames of this graph are in Figure 5.

4. EXPERIMENTAL RESULTS

We quantitatively evaluate the proposed fusion method on wide area imagery [12] and standard video sequences [2, 14, 1]. The wide-area data set has some challenging conditions such as shade occlusion, turning vehicles and low contrast; appearance changes are typical to standard sequences. We set up a tracking test-bed system as shown Figure 3 and run the tracking system with equal weight [9], variance ratio [16], distractor index [13] and feature prominence fusion schemes. We first calculate the weights of features from the previous frame. Then we use these weights to fuse the likelihood maps in the next frame. Except for the weighting scheme, all other testing parameters are equal. A better fusion approach would produce better target localization and improves the tracking performance.

The test tracking system determine the local maximum of the fused likelihood map as the target location. Therefore, we measure the local maxima of the fused maps of each method. Better fusion scheme should localize the highest peak location as the target. Figure 10 and Figure 11 show the peak probabilities of two sequences in aerial data set. According to the peak probabilities, the proposed fusion approach produces higher peak at the target location which increases the performance of the target localization.

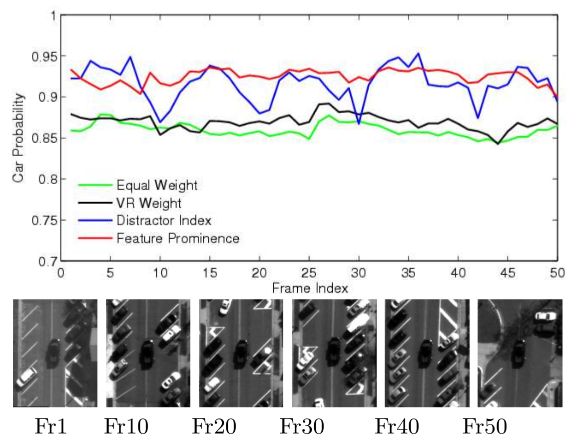

Figure 10: Top: Peak probability graphs of a wide area image sequence. Bottom: Image chips of sequences. There is no occlusion in these frames.

Figure 12 shows performance based on average recall versus number of selected peaks for each fusion method over all frames in the test sequences (15 cars with complex appearance and background environments). Each car has 100 or more frames in the video. It can be seen that the proposed feature prominence-based fusion approach gives the best recall over the other fusion approaches. The difference in performance between Feature Prominence and Distractor Index decreases until the two reach a similar performance starting around ninth peaks.

The peak probability scores and average peak-recall graph indicates that feature prominence method improves the target localization. In order to show that improved target localization increase the tracking performance, we measure the tracking performance of the fusion methods. The graph in Figure 13 indicates the precision-recall values of fusion

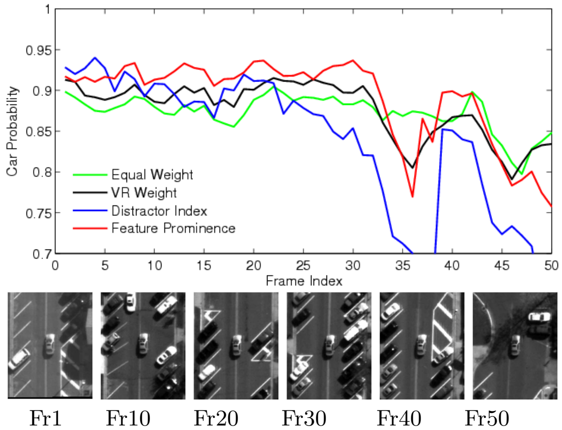

Figure 11: Top: Peak probability graphs of a wide area image sequence. Bottom: Image chips of sequences. There is no occlusion in these frames.

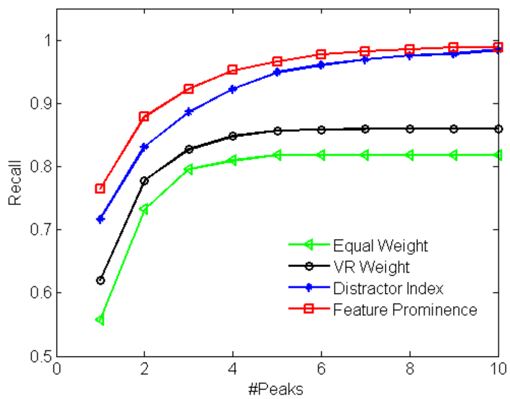

Figure 12: Average recall versus number of peaks selected for each fusion method using all frames, averaged across the fifteen test cars.

methods on four standard sequences each have 500 or more frames. According to the graph, feature prominence method produces better tracking results.

We also measure the position distance of the tracked locations of each fusion method. Figure 14 shows the position distance of fusion methods on the standard sequences. Distractor index and Feature prominence based methods produces better results than variance ratio and equal weighting fusion methods. Table 1 summarizes the average position distances, and Figure 15 gives example tracking results on some selected frames.

5. CONCLUSIONS

We propose *feature prominence* as a new feature fusion methodology that can be used to combine evidence from multiple feature operators for more reliable foreground target detection and localization during video tracking. Feature prominence is measured using a statistical significance measure. We combined the features according to their promi-

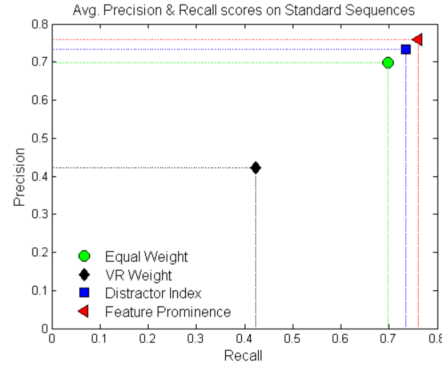

Figure 13: Average precision and recall scores for four standard data set including David (460 Fr), Faceocc (883 Fr), Faceoccc2 (810 Fr) and Girl(500 Fr).

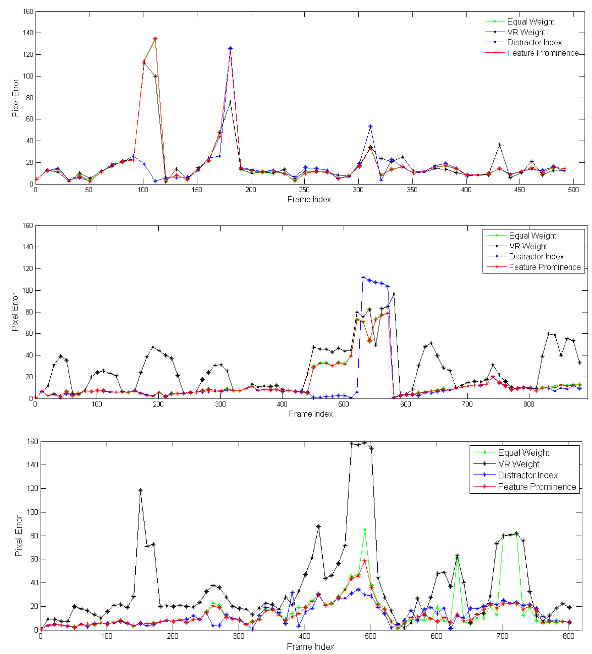

Figure 14: Position Distance Errors. Top: Girl Seq. Middle: Faceocc Seq. Bottom: Faceocc2 Seq.

Sequence	Fusion Methods E-W	VR-W	DI	FP
Girl	18.5	21.72	**14.95**	**15.95**
Faceocc	14.06	27.75	**10.86**	**10.25**
Faceocc2	15.73	36.28	**12.32**	**12.48**
David	56.48	63.81	60.45	**44.62**
Avg	26.21	37.39	24.65	**20.83**

Table 1: Mean Position Errors of Fusion Methods on Standard Videos. E-W: Equal Weight, VR-W: Variance Ratio Weight, DI: Distractor Index, FP: Feature Prominence.

Figure 15: Tracking results of Girl, Faceocc and Faceocc2 sequences. Black bounding box (bb) is Ground Truth, Green bb. - Equal Weight, Yellow bb. - Variance Ratio Weighting, Blue bb. - Distractor Index Weighting, Red bb. - Feature Prominence weighting schemes.

nence based weights. Preliminary results show that the new fusion approach is more robust than several other adaptively updated weighting schemes proposed in the literature. It does not underweight or overweight the features as much as the Variance Ratio or Distractor Index methods do. Moreover, it evaluates the peak probability with respect to the global distribution instead of local regional likelihoods as in the Variance Ratio. This makes the target probability estimation more robust within the search window.

6. ACKNOWLEDGMENTS

This research was partially supported by U.S. Air Force Research Laboratory (AFRL) under agreement AFRL FA8750-11-C- 0091. Approved for public release (case 88ABW-2012-0927). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of AFRL or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints.

7. REFERENCES

[1] A. Adam, E. Rivlin, and I. Shimshoni. Robust fragments-based tracking using the integral histogram. In *IEEE CVPR*, pages 798–805, 2006.

[2] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *IEEE CVPR*, pages 232–237, 1998.

[3] F. Bunyak and *et al.* Geodesic active contour based fusion of visible and infrared video for persistent object tracking. In *IEEE Workshop App. of Computer Vision*, 2007.

[4] S. Candemir and Y. S. Akgul. Statistical significance based graph cut regularization for medical image segmentation. *Turk. J. Elec. Eng. Comp. Sci.*, 19(6):957–972, 2011.

[5] S. Candemir and Y. S. Akgul. Statistical significance based graph cut segmentation for shrinking bias. In *Proc. Int. Conf. Image Analysis Recognition*, pages 304–313, 2011.

[6] R. Chellappa, A. Sankaranarayanan, A. Veeraraghavan, and P. Turaga. Statistical methods and models for video-based tracking, modeling, and recognition. *Found. Trends Signal Process.*, 3:1–151, 2010.

[7] R. T. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *IEEE Trans. PAMI*, 27:1631–1643, 2005.

[8] S. J. Julier and *et al.* The challenge of scalable and distributed fusion of disparate sources of information. In *SPIE Multisensor, Multisource Information Fusion*, volume 6242, 2006.

[9] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. PAMI*, 20(3):226–239, 1998.

[10] E. Maggio and A. Cavallaro. *Video Tracking: Theory and Practice*. Wiley, 2011.

[11] E. Maggio, F. Smeraldi, and A. Cavallaro. Adaptive multi-feature tracking in a particle filtering framework. *IEEE Trans. on Circuits and Systems for Video Technology*, 17:1348–1359, 2007.

[12] K. Palaniappan, R. Rao, and G. Seetharaman. Wide-area persistent airborne video: Architecture and challenges. *Distributed Video Sensor Networks: Research Challenges and Future Directions*, 2010.

[13] K. Palaniappan and *et al.* Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video. In *13th Conf. Information Fusion*, pages 1–8, 2010.

[14] D. Ross, J. Lim, R. Lin, and M. Yang. Incremental learning for robust visual tracking. In *IJCV*, pages 125–141, 2008.

[15] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. volume 38, 2006.

[16] Z. Yin, F. Porikli, and R. T. Collins. Likelihood map fusion for visual object tracking. In *IEEE Workshop Appl. Comput. Vis.*, pages 1–7, 2008.