PARTICLE PHD FILTERING FOR MULTI-TARGET VISUAL TRACKING

Emilio Maggio [1], *Elisa Piccardo* [1,2], *Carlo Regazzoni* [2], *Andrea Cavallaro* [1*]

[1] Multimedia and Vision Group - Queen Mary, University of London, United Kingdom
[2] DIBE - University of Genova, Italy
Email: emilio.maggio@elec.qmul.ac.uk; carlo@dibe.unige.it; andrea.cavallaro@elec.qmul.ac.uk

ABSTRACT

We propose a multi-target tracking algorithm based on the Probability Hypothesis Density (PHD) filter and data association using graph matching. The PHD filter is used to compensate for miss-detections and to remove noise and clutter. This filter propagates the first order moment of the multi-target posterior (instead of the full posterior) to reduce the growth in complexity with the number of targets from exponential to linear. Next the filtered states are associated using graph matching. Experimental results on face, people and vehicle tracking show that the proposed multi-target tracking algorithm improves the accuracy of the tracker, especially in cluttered scenes.

Index Terms— Multi-target, tracking, PHD filter, Monte Carlo methods, clutter.

1. INTRODUCTION

One of the major challenges in multi-target tracking is the estimation of the number of targets and their position in the scene, based on a set of uncertain observations. Uncertainty is due to noisy observations, occlusions, and clutter.

The problem of noise affecting state estimation is usually addressed in single-target tracking by Kalman and Particle filters [1]. The extension of these algorithms to multi-target tracking is difficult as the dimensionality of the state space is time variant. A possible solution is to track each target separately by independently applying single-target algorithms [2]. Another solution is to predefine a maximum number of targets (i.e., the dimensionality) and then to declare a group of targets as hidden [3]. A more elegant solution is to consider the multi-target set as a single meta target [4] and the observations as a single set of measurements of the meta sensor [5]. In this case, the multi-target state can be represented by a Random Finite Set (RFS), whose Bayesian propagation is similar to the single target case. However, the dimensionality of the target state still grows exponentially with the number of targets. Also the approximation of the RFS with Monte Carlo sampling requires a number of samples that grows exponentially, thus making the propagation of the full posterior unpractical. A less computationally intensive alternative is to propagate the Probability Hypothesis Density (PHD) (i.e., the first moment of the multi-target posterior)([4]). The integrals of the PHD recursion can be approximated with the samples generated by a Sequential Monte Carlo (SMC) method ([5]). As the PHD has the dimensionality of the single-target state, efficient sampling requires a number of particles that is proportional to the expected number of targets, thus leading to linear complexity. The cost for the lower complexity is that the PHD does not provide any information on the identity of the targets. Filtering techniques based on the PHD have been tested on synthetic data [5, 6], 3D sonar data [7], feature point filtering [8], and groups of humans detection [9]. As no data association is performed [5, 6, 8, 9] nor the target size is estimated [7], none of the above implementations can be applied to multi-target visual tracking.

In this paper we propose a multi-target tracker based on PHD filtering and graph matching for video object tracking in real-world scenarios. The PHD filter is adapted to reduce clutter and noise generated by common object detectors available for scene analysis. Next, a modified K-means clustering that accounts for the SMC sampling of the PHD is used to detect the peaks. Finally, the centers of the clusters are the input of the data association algorithm based on the maximum path cover of a bi-partitioned graph. The algorithm is demonstrated on real world scenarios with two object detectors, namely a motion detector and a face detector.

The paper is organized as follows. Sec. 2 introduces the Particle PHD filter. Sec. 3 describes the clustering and data association methods. In Sec. 4 we show the results on surveillance and face tracking scenarios. Finally in Sec. 5 we draw the conclusions.

2. THE PARTICLE PHD FILTER

Let us approximate the target area in an image with a $w \times h$ rectangle centered at $\mathbf{y} = (y_1, y_2)$. Let the state of a single target at time k be $x_k = (\mathbf{y}, \mathbf{v}, w, h) \in E_s$, where $\mathbf{v} = (v_1, v_2)$ is the target speed and E_s is the single target space. Finally, let the single target observation $z_k = (\mathbf{y}, w, h) \in E_o$ be generated by an object detector (e.g., a face detector or a motion detector). The corresponding multi-target state X_k and measurement Z_k are the finite collection of the states and measurements of each target. If $M(k)$ is the number of targets in the scene at time k, then $X_k = \left\{x_{k,1}, ...x_{k,M(k)}\right\}$ is the multi-target state. $Z_k = \left\{z_{k,1}, ...z_{k,N(k)}\right\}$ is the multi-target measurement formed by the $N(k)$ observations, where some of these observations may be due to clutter.

The uncertainty in the state and measurement is introduced by modeling multi-target state and multi-target measurement with two Random Finite Sets (RFS) Ξ_k and Σ_k. Ξ_k includes the information related to multi-target interactions, single target motion, and appearing or disappearing targets [4]. Similar to the single target case, the dynamics of Ξ_k is described by the multi-target transition density $f_{k|k-1}(X_k|X_{k-1})$, while Σ_k is described by the multi-target likelihood $g_k(Z_k|X_k)$.

The Probability Hypothesis Density (PHD) $D_{k|k}$ is the first order moment of a RFS, defined as the density $D_{k|k}(x)$ whose integral on any region S of the state space gives the expected number of targets in S. The PHD is a function in the single-target state space whose peaks identify the likely position of the targets. The Bayesian

*The authors acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC), under grant EP/D033772/1

iterative prediction and update of $D_{k|k}$ is known as the PHD filter.

The prediction operator, $(\Phi_{k|k-1}\alpha)(x)$, is defined as

$$(\Phi_{k|k-1}\alpha)(x) = \int \phi_{k|k-1}(x,\zeta)\alpha(\zeta)\lambda(d\zeta) + \gamma_k(x), \quad (1)$$

where α is any integrable function on E_s; $\gamma_k(x)$ is the intensity function of the new target birth RFS (i.e., the integral of $\gamma_k(x)$ over a region S approximates the average number of new objects per frame appearing in S); and $\phi_{k|k-1}(x,\xi)$ is the analogue of the state transition probability in the single target case:

$$\phi_{k|k-1}(x,\xi) = e_{k|k-1}(\xi)f_{k|k-1}(x|\xi) + \beta_{k|k-1}(x|\xi), \quad (2)$$

where $e_{k|k-1}$ is the probability that the target still exists at time k, $\beta_{k|k-1}(x|\xi)$ is the intensity of the RFS that a target is spawned from the state ξ, and $f_{k|k-1}(.|.)$ is the single target transition probability.

The update operator, $(\Psi_k\alpha)(x)$, is defined as

$$(\Psi_k\alpha)(x) = \left[p_M(x) + \sum_{z\in Z_k} \frac{\psi_{k,z}(x)}{\kappa_k(z) + \langle\psi_{k,z},\alpha\rangle} \right] \alpha(x), \quad (3)$$

where $p_M(x)$ is the miss-detection probability; $\psi_{k,z}(x) = (1 - p_M(x))g_k(z|x)$, and $g_k(z|x)$ is the single target likelihood defining the probability that z is generated by a target with state x; $\langle f,g\rangle = \int f(x)g(x)dx$, and $\kappa_k(z)$ is the clutter intensity function.

The PHD filter combines the single target probabilities with the other densities that model the interactions between multiple targets. A numerical solution of the integrals is obtained using a SMC method that approximates the PHD with a large set of weighted random samples (particles). Let the set $\{x_{k-1}^{(i)}\}_{i=1\ldots L_{k-1}}$ of L_{k-1} particles and associated weights $\{\omega_{k-1}^{(i)}\}_{i=1\ldots L_{k-1}}$ approximate the PHD at time $k-1$. By substituting the Monte Carlo approximation of the PHD (i.e., a finite sum of Dirac's deltas) in Eq. (1) and Eq. (3), and developing the equations [5], the following procedure is derived. First a new set $\{\tilde{x}_k^{(i)}\}_{i=1\ldots L_{k-1}+J_k}$ is generated by drawing L_{k-1} samples from the importance function $q_k(.|x_{k-1}^{(i)}, Z_k)$; these samples propagate the tracking hypotheses from the samples at time $k-1$. Then J_k samples are drawn from the new born importance function $p_k(.|Z_k)$, representing the state hypotheses of new targets appearing in the scene. The predicted weights, $\tilde{\omega}_{k|k-1}^{(i)}$, are defined as

$$\tilde{\omega}_{k|k-1}^{(i)} = \begin{cases} \frac{\phi_k\left(\tilde{x}_k^{(i)},\tilde{x}_{k-1}^{(i)}\right)\omega_{k-1}^{(i)}}{q_k\left(\tilde{x}_k^{(i)}|\tilde{x}_{k-1}^{(i)},Z_k\right)} & i=1,...,L_{k-1} \\ \frac{\gamma_k(\tilde{x}_k^{(i)})}{J_k p_k\left(\tilde{x}_k^{(i)}|Z_k\right)} & i = L_{k-1}+1,\ldots,L_{k-1}+J_k \end{cases}. \quad (4)$$

Once the new set of observations is available, the weights $\{\tilde{\omega}_{k|k-1}^{(i)}\}_{i=1\ldots L_{k-1}+J_k}$ are updated according to

$$\tilde{\omega}_k^{(i)} = \left[p_M(\tilde{x}_k^{(i)}) + \sum_{z\in Z_k} \frac{\psi_{k,z}(\tilde{x}_k^{(i)})}{\kappa_k(z) + C_k(z)} \right] \tilde{\omega}_{k|k-1}^{(i)}, \quad (5)$$

where $C_k(z) = \sum_{j=1}^{L_{k-1}+J_k} \psi_{k,z}(\tilde{x}_k^{(i)})\omega_{k|k-1}^{(j)}$.

At each iteration, J_k new particles are added to the old ones. To limit the growth of the number of particles, L_k particles are resampled from $\left\{\tilde{\omega}_k^{(i)}/\hat{M}_{k|k}, \tilde{x}_k^{(i)}\right\}_{i=1}^{L_k+J_k}$, where $\hat{M}_{k|k}$ is the total mass. L_k is chosen to keep the number of particles per target, ρ, constant. At each time step, a new L_k is computed so that $L_k = \rho\hat{M}_{k|k}$.

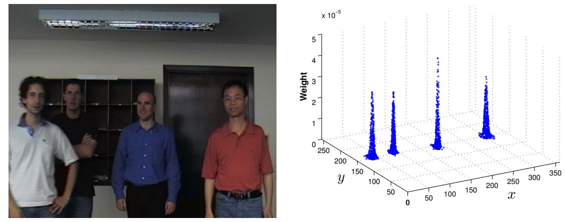


Fig. 1. Visualization of the particles approximating the PHD on the frame at the left when the faces are the targets.

Hence the complexity of the algorithm grows *linearly* with the number of targets in the scene. After resampling, the weights of set $\left\{\omega_k^{(i)}, x_k^{(i)}\right\}_{i=1}^{L_k}$ are normalized to preserve the total mass.

In our case, the original particle PHD filter ([5]) has to be adapted to account for the two additional dimensions corresponding to w and h. We define the state transition $f_{k|k-1}(x|\xi)$ assuming that each target moves according to a first-order Gaussian dynamic on the position and zero-order on the size. As larger objects (in the image plane) usually accelerate faster than smaller objects in the same scene, the intensity of the white Gaussian noise modeling acceleration and change of size is proportional to the size of the target at time $k-1$. For simplicity no spawning of targets is considered. The single-target likelihood $g_k(z|x)$ is a Gaussian $\mathcal{N}(z, C_z)$ centered on the observation. The covariance matrix C_z is diagonal with standard deviations proportional to the sizes w and h of each $z_{k,i}$; the miss-detection probability p_M is uniform. The L_{k-1} old particles are propagated according to the dynamics, i.e. $q_k(.|.) \propto f_{k|k-1}(.|.)$. While J_k new born particles are drawn from a mixture of Gaussians centered on the observations Z_k. The $N(k)$ coefficients of the mixture are set computing the cumulative masses $M(z_{k,j})$ of the propagated particles that are inside the 99% confidence interval determined by the observation noise. Then, $J_{z_{k,j}}$ new born particles are allocated to each detection according to $J_{z_{k,j}} = J_{max}(1 - M(z_{k,j}))$, and $J_k = \sum_j J_{z_{k,j}}$. The birth intensity $\gamma_k(x)$ is uniform around the parameters $\mathbf{y}$, w and h, uniform on the velocity direction and Gaussian on the amplitude.

3. CLUSTERING AND DATA ASSOCIATION

The PHD is represented by a set of particles $\{x_k^{(i)}\}_{i=0\ldots L_k}$ defined in the single-target state space. An example of PHD approximated by particles is shown in Fig. 1. The peaks of the PHD are on the detected faces, and the mass $\hat{M}_{k|k} \approx 4$ estimates the number of targets. The weights of the particles are higher where the tracking hypotheses are validated by consecutive detections. A clustering algorithm is now required to detect the peaks of the PHD and then data association will be applied on the cluster centers to perform tracking.

Particles are clustered with *K-means*, using a special procedure to select the number of clusters. In fact, although $\hat{M}_{k|k}$ estimates the number of targets, it is not a good estimator of the number of clusters, especially when targets are appearing or disappearing and their total mass is smaller than 1. $N(k)$ clusters are first initialized on the $N(k)$ observations with zero velocity. After convergence of K-means, if $M_k^{(j)} > 1$ (i.e., the total mass of the samples in cluster j is larger than 1) a new center is set on the cluster sample with largest distance from the center. Next, K-means iterates until the condition $M_k^{(j)} < 1$, $\forall j$ is satisfied. Finally, clusters with $M_k^{(j)} < T$ are discarded, where T is a predefined threshold. The remaining centers

Fig. 2. Examples of Particle PHD filtering (Filtered states) which removes cluttered detections (Detections), for example as in column (b,d,e). Comparison of tracking results between the multi-target tracker with Particle PHD filtering (PHD-MT) and without (MT). The tracks are color coded. As showed in column (e) and (f) PHD-MT successfully recovers the faces after a total occlusion without generating false tracks.

$\hat{X}_k$ are selected as the output set of states used for data association. The information contained in $\hat{X}_k$ includes an estimate of the target velocity and is therefore richer than that from the observations.

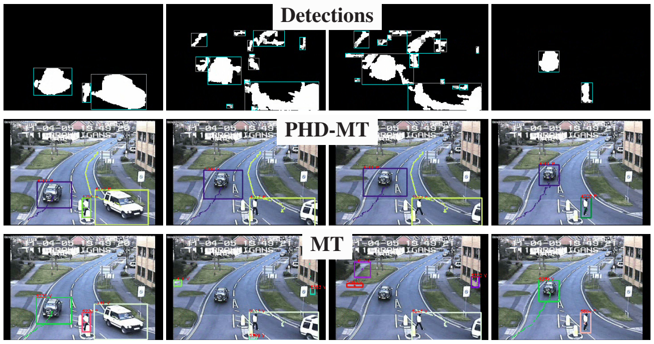
To obtain a consistent identity of a target over time, the cluster centers are used as vertices in a *graph matching* procedure ([10]), which is used for data association. Let a cluster center $\hat{x}_k \in \hat{X}_k$ be represented by a vertex $v(\hat{x}_k) \in V_k$ of the graph G, where V_k is the set of vertices representing the targets at time k. The tentative associations between candidate targets at different time instants are described by the gain associated with each edge in G. The graph is formed by iteratively creating new edges from the old set of vertices $\{V_{k-j}\}_{j=1\ldots K}$ to the new set of vertices V_k associated to cluster centers of frame k. The possible edge combinations represent multiple track hypotheses, including miss-detections and occlusions (i.e., edges between two vertices $v(\hat{x}_k)$ and $v(\hat{x}_{k-j})$, with $j > 1$). The path cover of G with the maximum gain identifies the best set of tracks. We define the gain of the edge by $log(f(\hat{x}_k|\hat{x}_{k-j}))$. In this case the optimal path cover maximizes the likelihood over the possible sets of tracks represented by the edges in the graph. After maximization, a *new target* is modeled by a vertex without backward correspondence, whereas a *disappeared target* is modeled by a vertex without forward correspondence. The maximum number of consecutive occluded frames during which an object track can be recovered is defined by the graph depth K, as discussed next.

4. EXPERIMENTAL RESULTS

We demonstrate the proposed tracker on real-world test sequences with faces, people and vehicles. The face sequence is available at http://www.elec.qmul.ac.uk/staffinfo/andrea/PHD-MT.html, while people and vehicle sequences are from the VACE dataset [11]. Face detections are generated with an Adaboost classifier [12], and people and vehicle detections with a color based change detector [13]. The sequences are converted to QCIF format and 25 Hz before processing. *The parameters used in the simulations are the same for all test sequences*. The multi-target tracker based on the particle PHD filter (MT-PHD) is compared with the multi-target tracker (MT) where the data association described in Sec. 3 is performed directly on Z_k.

The parameters of the tracker are described in the following. The particle PHD filter uses $\rho = 1500$ particles per target and $J_{max} = 1000$ particles per detection. The standard deviations of the transition model are: $\sigma_{k,v_1}(w_{k-1}) = 0.03w_{k-1}$; $\sigma_{k,v_2}(h_{k-1}) = 0.03h_{k-1}$; $\sigma_{k,w}(w_{k-1}) = 0.1w_{k-1}$; $\sigma_{k,h}(h_{k-1}) = 0.1h_{k-1}$. The standard deviations of the Gaussian observation noise are: $\sigma^z_{k,w}(w_{k-1}) = \max(0.15w_{k-1}, 1.0)$; $\sigma^z_{k,h}(h_{k-1}) = \max(0.15h_{k-1}, 1.0)$; $\sigma^z_{k,y_1} = \sigma^z_{k,w}/2$; $\sigma^z_{k,y_2} = \sigma^z_{k,h}/2$. The birth intensity is 0.01 targets per frame per detection. The intensity of clutter around the detections is 1.0 clutter point per frame. $T = 0.6$ is used to accept the cluster centers as real targets. The miss-detection probability $P_M = 0.05$, and the survival probability $e_{k|k-1} = 0.99$. For the data association the depth of the graph is $K = 50$ frames, hence the algorithm is capable of resolving occlusions for a maximum of 2 seconds on a 25 Hz footage. Tracks shorter than 5 frames are discarded.

Fig. 2 shows the results of the PHD filter on the face detections. When false detections are processed (Fig. 2 (a)(b)(d)(e)(f)) the mass of the PHD starts growing around them; however, multiple coherent and consecutive detections are necessary to increase the mass to a level greater than T. When the clutter is not persistent, the PHD filter successfully manages to remove it. The drawback of the filtering is a slower response in detecting new real targets due to the trade-



(a) (b) (c) (d)

Fig. 3. Comparison of tracking results between the multi-target tracker with Particle PHD filtering (PHD-MT) and without (MT). False tracks due to clutter ((b) and (c)) are removed by PHD-MT.

off between clutter removal and response time. The particle PHD also manages to smooth the errors in estimating the correct size of the face (Fig. 2(c)(d)). Furthermore, Fig. 2 shows how the combination of PHD filtering with the graph-based data association is able to recover the faces after a total occlusion: in Fig. 2 (e)(f) a face is occluded by the other person, and data association successfully links the tracks. Finally, the results of PHD-MT compared with MT shows that false tracks due to clutter are removed by PHD-MT only (Fig. 2 (e)(f)). Similar considerations can be done on the PHD-MT used to process the output of the change detector (Fig. 3). In this challenging situation, generated by a sudden change in illumination, although the accuracy in recovering the size of the target is poor, the heavy clutter is filtered by PHD-MT (Fig. 3(b)(c)). Furthermore, the smoothing generated by the filtering allows the data association to connect the target states.

Table 1 shows the comparative results of PHD-MT and MT based on the VACE protocol [11], which is composed of four scores, namely Multiple Object Detection Accuracy (MODA), Detection Precision (MODP), Tracking Accuracy (MOTA), and Tracking Precision (MOTP). The evaluation based on the scores is performed on sequences where ground-truth data are available. It is possible to notice that PHD-MT outperforms MT for all the scores and sequences. The improvement is larger when the tracking system is challenged by clutter, that generates false detections to be removed (i.e. *Seq102a03* and *Seq201c01*). In all sequences PHD-MT is more effective in removing clutter than MT thus improving both precision and accuracy.

In terms of computational complexity the Particle PHD filter uses 34% of the total computational power of the overall tracking algorithm when used with the change detector, whereas clustering and data association account for 2% only. The remaining 64% of the time is spent on the change detector. When the face detector is used, the percentages are 78% for the detector and 20% for the PHD filter. The full *non-optimized* tracking algorithm runs at an average of 6fps on a processor Pentium IV 3GHz. We expect that the overall algorithm can achieve real-time performance with code optimization.

5. CONCLUSIONS

We presented a multi-target visual tracker that employs Particle PHD filtering to remove clutter and miss-detections from noisy observations. The resulting set of particles is clustered by a modified K-means adapted to the Particle PHD. To generate the final tracks the

Table 1. Comparison of tracking accuracy results between the multi target tracker with (PHD-MT)(1) and without (MT)(2) PHD filter.

Seq.	MODP		MODA		MOTP		MOTA	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
102a01	0.61	0.58	0.54	0.49	0.65	0.60	0.53	0.48
102a03	0.60	0.59	0.16	-0.58	0.60	0.57	0.15	-0.58
102a11	0.59	0.58	0.28	0.27	0.59	0.58	0.28	0.27
201c01	0.37	0.36	0.44	0.36	0.39	0.37	0.43	0.35

centers of the clusters are processed by a data association algorithm based on graph matching. The proposed algorithm has the capability to remove non persistent clutter, filter miss-detections, smooth the tracks, and overcome short term occlusions. Experimental results over a set of real-world sequences show that the Particle PHD Filter improves the robustness of the tracker to clutter by verifying the coherence of consecutive sets of detections, without significantly increasing the complexity of the overall algorithm. Future work includes the definition of a model of the merging and splitting of the tracks that combines the capabilities of the PHD filter, with the information contained in the vertices of the graph.

6. REFERENCES

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[2] K. Okuma, A. Taleghani, N. De Freitas, J.J. Little, and D. G. Lowe, “A boosted particle filter: Multitarget detection and tracking,” in *Proc. of the European Conf. on Computer Vision*, Prague, CZ, May 2004, vol. 1, pp. 28–39.

[3] L.D. Stone, *A Bayesian approach to multiple-target tracking*, Handbook of Multisensor Data Fusion. CRC Press, June 2001.

[4] R. Mahler, “A theoretical foundation for the Stein-Winter probability hypothesis density multitarget tracking approach,” in *Proc. MSS Nat’l Symp. on Sensor and Data Fusion*, San Antonio, TX, June 2002.

[5] B.N. Vo, S.R. Singh, and A. Doucet, “Sequential monte carlo implementation of the PHD filter for multi-target tracking,” in *Proc. of Intern. Conf. of Information Fusion*, Cairns, AU, July 2003, pp. 792–799.

[6] H. Sidenbladh and S. Wirkander, “Tracking random sets of vehicles in terrain,” in *Proc. of IEEE Workshop on Multi-Object Tracking*, Madison, WS, USA, June 2003.

[7] D.E. Clark and J. Bell, “Bayesian multiple target tracking in forward scan sonar images using the PHD filter,” *IEE Proc. Radar, Sonar and Navigation*, vol. 152, pp. 327–334, Oct. 2005.

[8] H. Maeda N. Ikoma, T. Uchino, “Tracking of feature points in image sequence by smc implementation of phd filter,” in *Proc. of SICE 2004 annual*, Aug. 2004, vol. 2, pp. 1696–1701.

[9] Y.D. Wang, J.K. Wu, A.A. Kassim, and W.M. Huang, “Tracking a variable number of human groups in video using probability hypothesis density,” in *Proc. of IEEE International Conf. on Pattern Recognition*, Hong Kong, CHN, Aug. 2006, vol. 3, pp. 1127–1130.

[10] K. Shafique and M. Shah, “A noniterative greedy algorithm for multiframe point correspondence,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 1, pp. 51–65, 2005.

[11] R. Kasturi, “Performance evaluation protocol for face, person and vehicle detection & tracking in video analysis and content extraction,” *Computer Science and Engineering, University of South Florida*, Jan. 2006, http://isl.ira.uka.de/clear06/downloads/ClearEval_Protocol_v5.pdf.

[12] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. of IEEE Conf. on Comp. Vis. and Pattern Recog.*, Kauai, HI, USA, 2001, vol. 1, pp. I–511–I–518 vol.1.

[13] A. Cavallaro and T. Ebrahimi, “Interaction between high-level and low-level image analysis for semantic video object extraction,” *EURASIP Journal on Applied Signal Processing*, vol. 6, pp. 786–797, June 2004.