Autonomous Detection and Tracking under Illumination Changes, Occlusions and Moving Camera.

Harish Bhaskar[a,b], Kartik Dwivedi[c], Debi Prosad Dogra[d], Mohammed Al-Mualla[a], Lyudmila Mihaylova[e]

[a]*Khalifa University of Science, Technology and Research (KUSTAR), P.O.Box 127788, Abu Dhabi, U.A.E.*
[b]*Merchant Venturers' Building, Woodland Road, Clifton, Bristol BS8 1UB, U.K.*
[c]*Indian Institute of Technology, Guwahati*
[d]*Indian Institute of Technology, Bhubaneswar*
[e]*Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD*

Abstract

In this paper, an autonomous multiple target detection and tracking technique for dynamic scenes that are influenced by illumination variations, occlusions and camera instability is proposed. The framework combines a novel Dynamic Reverse Analysis (DRA) approach with an Enhanced Rao-Blackwellized Particle filter (E-RBPF) for multiple target detection and tracking respectively. The DRA method, in addition to providing accurate target localization, presents the E-RBPF scheme with costs associated with the differences in intensity caused by illumination variations between consecutive frame pairs in any video of a dynamic scene. The E-RBPF inherently models these costs, thus allowing the framework to a) adapt learning parameters, b) distinguish between camera-motion and object-motion, c) deal with sample degeneracy, d) provide appropriate appearance compensation during likelihood measurement and e) handle occlusion. The proposed detect-and-track method when compared against other competing baseline techniques has demonstrated superior performance both in accuracy and robustness on challenging videos from publicly available datasets.

*Corresponding author
Email addresses: harish.bhaskar@kustar.ac.ae (Harish Bhaskar), harish.bhaskar@bristol.ac.uk (Harish Bhaskar), k.dwivedi@iitg.ac.in (Kartik Dwivedi), dpdogra@iitbbs.ac.in (Debi Prosad Dogra), l.s.mihaylova@sheffield.ac.uk (Lyudmila Mihaylova)

Keywords: Target Detection, Tracking, Illumination Variations, Occlusion, Camera Movements, Reverse Analysis, Rao-Blackwellized Particle Filter, Likelihood Model
2014 MSC: 00-01, 99-00

1. Introduction & Related Work

Detection and tracking are challenging research problems, particularly in unconstrained surveillance scenarios. Much research efforts have been spent in developing state-of-the-art detection and tracking methodologies including detect-and-track [1], track-before-detect [2, 3], Probability Hypothesis Density (PHD) filter based multiple target tracking techniques [4, 5], among many others. Despite advances, detection and tracking are still challenged by the presence of illumination variations [4], occlusions [6, 7], and camera movements [8]. Although many approaches have been proposed that address these issues in a mutually exclusive manner, the joint problem is still far from being solved.

The concept of illumination invariance during target detection and tracking has been addressed in different ways which can be categorized into feature-based [9] and appearance-based methods [10]. A good example of the feature-based detection technique can be found in [9], where a sparse set of salient illumination invariant features are considered. Similarly, in the work of [11, 12], the bi-parametrization of different combinations of color spaces have been studied for dynamic target tracking. However, such methods have failed to adequately discriminate targets against the background, during detection. A comprehensive overview of the recent efforts in background modelling based detection approaches can be found in [13, 14]. From a tracking point-of-view, some initial work has been done in joint target localization and estimation of illumination variations as in [15, 16]. Similarly, a method for coping with appearance changes of targets during tracking has been proposed in [17]. Although such methods are proven to handle constrained gradual illumination changes, modelling of illumi-

nation changes continues to be highly complex. This can mainly be attributed to factors such as: a) non-linearity of illumination changes in real-scenarios, b) ambiguity in the interpretation of the differences in intensity variations caused by the motion of targets as against due to illumination changes, and c) disregarding certain pertinent visual information to provide illumination invariance that can cause difficulties in handling occlusion and related challenges.

Occlusion detection during tracking is considered a hard problem in most general-purpose tracking algorithms [18]. The primary challenge in handling occlusion is to accumulate sufficient evidence from observations so that reliable data association becomes possible. Research indicates that performing occlusion handling within tracking is limited only to analysing pixel variations using multi-modal distributions in order to encompass statistical properties of occluders to distinguish it from the target(s)-of-interest [19]. However, most assumptions of feature-level similarity become invalid when considering real-world scenarios. In the study by [19], it has been shown that the contextual content which encapsulates motion information is also capable of handling occlusion. In another example, the problem of target initiation and termination has been shown to be handled using a hierarchical particle filtering framework [20]. Further, the use of spatio-temporal modelling has been proposed for human silhouette extraction from noisy and occluded data [7]. Though attempts have been made to tackle occlusion issues during tracking, the following complexities continue to remain: a) localization of the targets when occlusion is unknown, b) updating target descriptors during appearance changes, c) robustness against noise and clutter, and d) coping with disappearances and re-appearances of targets.

Motion in the background, target deformation and changes in the camera position during jitter all present similar effect on the spatio-appearance of targets during detection and tracking. In order to model spatio-temporal appearance changes of targets in the joint space, the use higher order distributions has become a popular choice [21]. Recent studies have focused on using Alpha-

stable [21] and Cauchy [22] distributions to model pixel intensity variations caused by camera shake during detection. Further, the generation of spatio-temporal methods for handling camera movements within a background modelling framework has also been recently proposed in [23]. However, the robustness of such models for dynamic scenes have not been fully explored.

This brief survey of the literature has clearly highlighted that the treatment of these challenges in a mutually exclusive manner cannot facilitate robust detection and tracking in real-world scenarios. On the other hand, incorporating different adaptations into a singular model may not always help solving all problems jointly. Therefore, in this paper, a tight integration of the constituent processes into a unified framework for the detection and tracking of dynamic scenes is proposed. It is hypothesized that an integrated detect-and-track technique capable of generating sufficient statistics of illumination variations during the accurate localization, when tunnelled across to an enhanced RBPF framework shall provide robust tracking of multiple targets within a dynamic scene.

2. Novelty & Contributions

One key novelty of the proposed framework is the use of DRA within background modelling for accurately detecting (or spatially localizing) multiple moving targets under changing illumination conditions. Furthermore, such a detection procedure allows extracting sufficient statistics that are indicative of the temporal location, type and extent of the illumination variations in the dynamic scene. Another important novelty is the tight integration of the qualitative and quantitative outputs from the DRA-based target detection method together with the E-RBPF tracking algorithm for adaptation against dynamic illumination and camera movements. Finally, a loose integration of the E-RBPF framework with appropriate noise models and likelihood measurements have allowed the framework to compensate for local (dis)order.

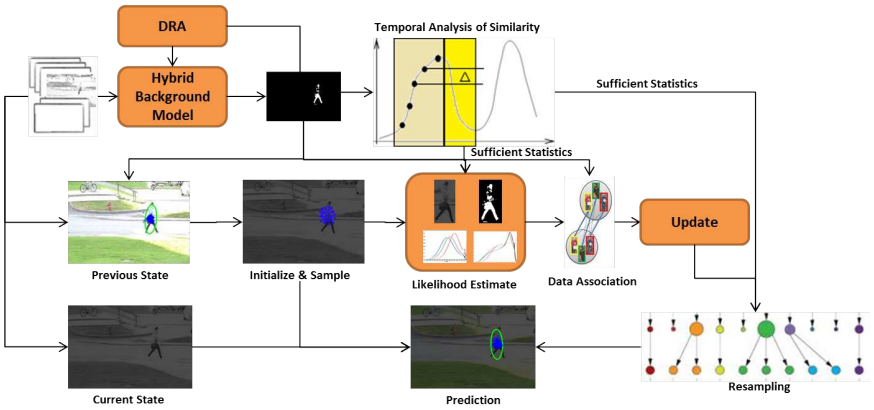


Figure 1: Functional block diagram of the proposed framework.

The rest of the paper is organized as follows. In Section 3, a detailed description of the proposed detect-and-track framework is presented. Following this, performance evaluation and comparison of the proposed detection and tracking methodologies against the state-of-the-art is described in Section 4. Section 5 concludes.

3. Proposed Methodology

The proposed framework for multiple target tracking is a modularized yet coupled approach. The method tightly integrates a) a hybrid background modelling scheme for accurate target detection with b) an enhanced particle filtering framework for robust target tracking. An illustration of the proposed framework is presented in Figure 1.

The proposed hybrid background modelling method combines conventional background initialization and maintenance processes with a reverse analysis scheme for accurate target detection. The DRA technique exploits deviations in the foreground detection procedure between forward and reverse directions to extract sufficient statistics on the changes in illumination conditions to determine a) the composition of optimal frames that produces a representative background

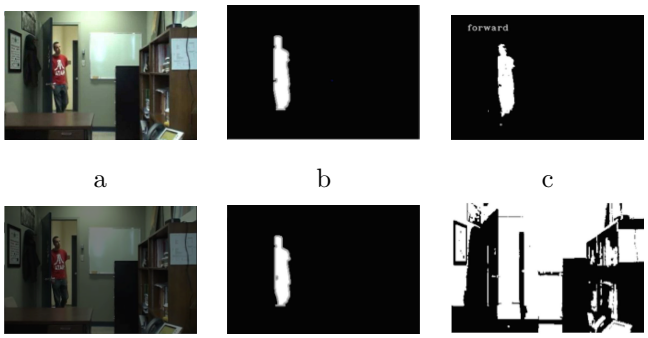


Figure 2: Illustration to demonstrate the impact of illuminations changes on background modelling. Results of background modelling on the original frame (a) and illumination changed frame (d) are presented in (c) and (f) respectively with the corresponding ground truth in (b) and (e).

model and b) control of adaptation parameters for coping with dynamic changes.

In contrast to the rigorous detection process, tracking is developed as an enhancement to the Rao-Blackwellized particle filter (E-RBPF) comprising of a data association and likelihood models by incorporating detection prior along with an advanced re-sampling scheme and spatio-temporal appearance separated noise model. The use of the detection prior within the likelihood model allows coping with occlusion and improving tracking accuracy while its use with the re-sampling method help solving the degeneracy problem.

3.1. DRA-based Background Modelling for Target Detection

Foreground (target) detection using background subtraction is based on the principle of updating an online statistical background model. Each pixel in a new image is classified as background if it fits the stochastic statistical background model, otherwise labelled foreground. During the process of building a statistical background model, the intensity variation of pixels over a history of previous frames is usually considered. The variations thus studied not only im-

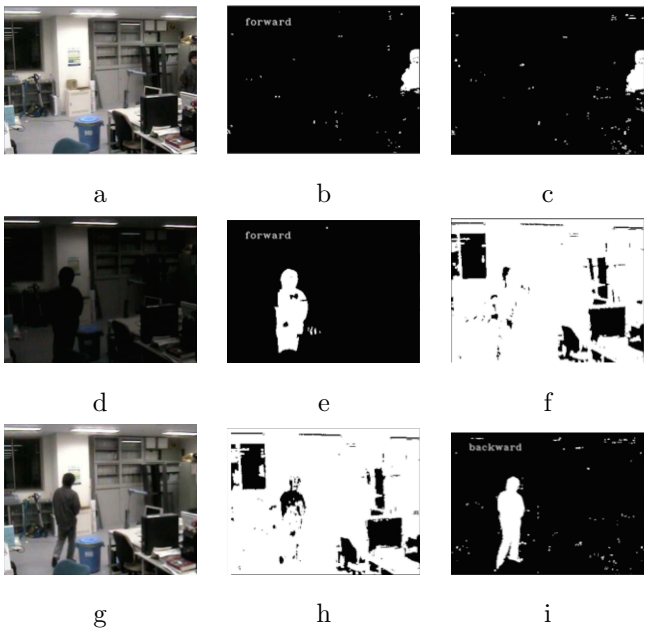


Figure 3: Results of forward background modelling in (b), (e) and (h) and backward background modelling (c), (f) and (i) on frames undergoing illumination changes as in (a), (d) and (g).

pact the correct classification of background pixels but also helps the background model adaptation during the maintenance phase. Therefore, a judicious selection of image frames for initializing the background model is key to achieving good detection accuracy, particularly under changing illumination conditions. Consider the example shown in Figure 2(a), acquired at constant illumination conditions and its corresponding labelled ground truth detection in Figure 2(b). When, a conventional adaptive background modelling technique, such as [24] is used for detection, it produces an output as in Figure 2(c). However, in the event of a sudden change in the illumination condition as in Figure 2(d) or Figure 2(g), it can be noticed that the foreground detection fails until further adaptation e.g. Figure 2(f) or Figure 2(i).

Reverse-time correlation (RTC) analysis is a popular method of tuning dynamics, popularly used for investigating model behaviour [25]. In order to

motivate the use of DRA in background modelling, Figure 3 is considered. The original images at different illumination conditions are displayed in Figure 3(a),Figure 3(d) and Figure 3(g), respectively. The results of foreground detection using a history of frames from the forward direction are presented in in Figure 3(b), Figure 3(e) and Figure 3(h). Similarly, the foreground prediction results using future frames in the reverse directions is as shown in Figure 3(c), Figure 3(f) and Figure 3(i). The disagreements in these results between the forward and reverse directions indicate changes in the illumination conditions of the environment.

In this paper, the disagreements between the forward and reverse prediction analysis of frames in foreground detection is the underlying motivation for the proposed DRA-based hybrid background modelling technique that is illustrated in Figure 4. The method works in two phases. In the first phase, a history of frames from the forward direction and future frames from the reverse direction are used independently to make their predictions of the current frame (frame 4 in Figure 4). A measure of similarity between the predictions of the forward and reverse directions is made and continued temporally. Further, a temporal analysis of the intensity variation allows extracting sufficient statistics of the illumination changes. These statistics enable determining the appropriate number of frames from the forward and reverse directions to be chosen to build an accurate hybrid background model to be used in the second phase. In addition, the adaptation (or learning) parameters of model are also updated using these statistics online during background maintenance.

A mathematical formulation of the DRA-based hybrid background modelling using GMM is described as follows. Here, each pixel is characterized by its intensity in a chosen color space (usually RGB color space). The posterior probability of observing the current pixel is formulated as:

$$P(X_t) = P(X_t|M^{(\overrightarrow{\xi_{t_1}},\overleftarrow{\psi_{t_2}})}) \tag{1}$$

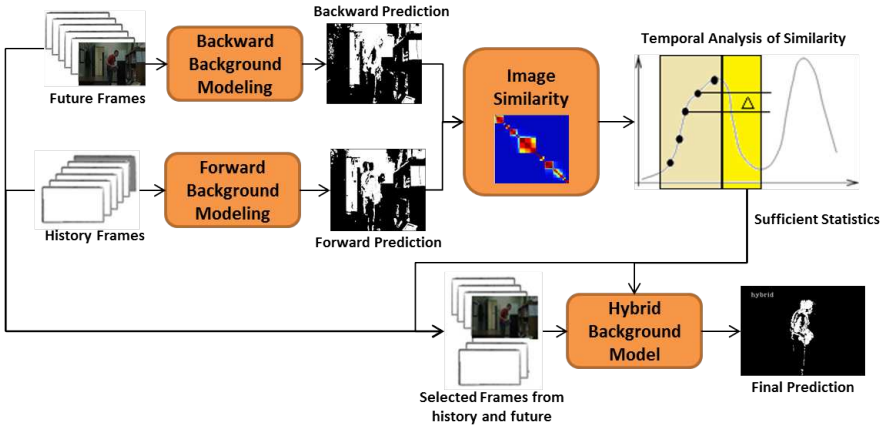


Figure 4: A process flow illustration of the proposed DRA-based Hybrid Background Modelling framework.

where X_t represents the intensity of pixels at time t, M represents a model of the background learnt from a selected subset of $\xrightarrow{\xi_{t_1}}$ image frames from the forward direction and $\xleftarrow{\psi_{t_2}}$ from the reverse direction.

$$P(X_t|M^{(\xrightarrow{\xi_{t_1}},\xleftarrow{\psi_{t_2}})} = \sum_{j=1}^{K} w_t^j \mathcal{N}(X_(t, \xrightarrow{\xi_{t_1}}, \xleftarrow{\psi_{t_2}}), \mu, \Sigma) \tag{2}$$

where w_t^j is the weight of the j^{th} Gaussian at time t. After initializing the mean (μ) and estimation of the covariance matrix (Σ) using an EM algorithm, foreground detection is performed. Next, the parameters are updated in order to maintain the background model. Once the K Gaussian variables are ordered appropriately, the first B Gaussian distributions which exceed certain threshold T_1 are retained for a background distribution:

$$B = \underset{b}{\operatorname{argmin}} \sum_{i=1}^{b} (w_t^j > T_1) \tag{3}$$

The other distributions are considered to represent the foreground (target). When a new image frame arrives at time instant $t+1$, the pixels are matched using a distance metric (usually Mahalanobis distance) to determine the matching Gaussian distribution and hence classify it as a foreground (target) or back-

ground. When a match is found to one of the K Gaussian variables, for the component matched, update is performed as:

$$\begin{aligned} w_{t+1}^{j} &= (1-\alpha)w_{t}^{j} + \alpha \\ \mu_{t+1}^{j} &= (1-\rho)\mu_{t}^{j} + \rho X_{(}t+1, \xrightarrow{\xi_{t_1}}, \xleftarrow{\psi_{t_2}}) \\ \sigma_{t+1}^{2,j} &= (1-\rho)\sigma_{t}^{2,j} + \rho(X_{t+1}^{j} - \mu_{t+1}^{j}).(X_{t+1}^{j} - \mu_{t+1}^{j})' \end{aligned} \tag{4}$$

where α and ρ are constant learning rates, conventionally referred to as the adaptation control parameter and convergence control parameter, respectively.

In lieu of above description of the problem, the main contributions of this paper are detailed below:

- determining the optimal composition of $\xrightarrow{\xi_{t_1}}$ frames in the forward direction and $\xleftarrow{\psi_{t_2}}$ frames in the reverse directions to be considered in building the model M of the pixel intensity variation X over time using a mixture of K Gaussian variables.
- adaptively estimating the values of the control parameters for adaptation α and convergence ρ by appropriately modelling the temporal changes in the illumination conditions.

The process of dynamic reverse analysis involves building directional foreground predictive models using the conventional GMM-based method using the posterior probability framework as below:

$$\begin{aligned} P_f(X_t) &= P(X_t|M^{(\xrightarrow{\epsilon_{t-1}})}) \\ P_b(X_t) &= P(X_t|M^{(\xleftarrow{\epsilon_{t+1}})}) \end{aligned} \tag{5}$$

where, $P_f(X_t)$ and $P_b(X_t)$ represent the posterior probabilities measured using the model $M^{(\xrightarrow{\epsilon_{t-1}})}$ in forward and $M^{(\xleftarrow{\epsilon_{t+1}})}$ in backward directions, respectively.

Further, a metric to evaluate the extent of similarity χ between foreground (target) predictions of the directional models is proposed. This similarity criteria χ consists of a weighted combination of: a) the difference between the foreground detected outputs of the current image frame I_t and a selected history of previous outputs $I_{t-k:t}$ for some constant k and b) the normalized amount of the foreground classified pixels in the current frame $\overline{I}_t^f$, where ϑ and υ are the weights.

$$\chi(P_f(.), P_b(.)) = \vartheta(d(I_t - I_{t-k:t})) + \upsilon(\overline{I}_t^f) \tag{6}$$

Finally, a temporal distribution of the similarity is modelled as $\chi_t(.)$. The peaks of this temporal distribution of similarity between the directional models is estimated using:

$$\zeta_n = \text{argmax}(\chi_t > T_2) \tag{7}$$

where, the n peaks ζ_n of this temporal distribution are analysed and landmarked to represent the points of change in the illumination conditions and T_2 is a predefined threshold. Sufficient statistics are estimated using such analysis to determine the class (type) of the illumination change, rate of change, between consecutive points by modelling the pixel intensity variations as the function that distinguishes this illumination change from the foreground (target) motion.

$$\kappa_n = \phi(X^n, X^{n+1}|\zeta_n) \tag{8}$$

Furthermore, a self-adaptive learning mechanism that parametrizes illumination changes using ϕ to model the pixel intensity variations producing κ. This parametrization process is used to approximate the composition of frames from both the forward and reverse directions through a localization function ℓ.

$$(\xi, \psi) = \ell(\kappa_n, X_t) \tag{9}$$

The localization function ℓ in conjunction with the mapping function κ and the location of the current frame with respect to n, allows computing approximate estimates for ξ and ψ together with the duration of illumination change measured as a distance, e.g. $\delta_n = d(\zeta_n, \zeta_{n+1})$ to allow building the hybrid background model and controlling the adaptation parameters α and ρ respectively, where d is any distance function, typically Euclidean distance.

3.2. E-RBPF based Target Tracking

In order to track multiple targets using the proposed E-RBPF technique, the target representation model is first chosen. In this paper, an 8-D ellipse model (10) is used to describe the dynamics of the target:

$$\{x, \Delta x, y, \Delta y, H, \Delta H, W, \Delta W\} \tag{10}$$

where (x, y) represents the centroid of the elliptical model, $(\Delta x, \Delta y)$ represents the velocity, (H, W) denotes the scale parameters - horizontal and vertical half lengths of the elliptical axes, and $(\Delta H, \Delta W)$ are the corresponding scale changes. This chosen model is standard and has been adopted in a manner similar to the work of [26]. The main use of the 8-D elliptical model is that it facilitates the splitting of the state space of the filter between root variables (R) containing motion information and leaf variables (L) consisting of the scale parameters. In general, the RBPF framework allows the propagation of the root variables one step ahead using (11).

$$R^t = \top R_{t-1} + \eta_{r,t-1} \tag{11}$$

where $\top$ represents the transition matrix and η_r is random noise. Given its conditional dependence on the root variables, the leaf variables form a linear-Gaussian substructure that is optimally estimated using a typical Kalman filter.

$$L^t = AL_{t-1} + \Phi(R_t, R_{t-1}) + \varepsilon_{t-1} \tag{12}$$

where Φ encodes the relationship between the leaf variable L and the root variables R from $t-1$ to t, ε is Gaussian random noise and A denotes a constant matrix. The image observations combining both the linear and non-linear states is represented as Z_t as given in (13).

$$Z_t = \Upsilon(R_t, L_t, \eta_{o,t}) \tag{13}$$

where η_o represents observation noise, Υ is a non-linear functional mapping and the auxiliary observations corresponding to the leaf variables that share a linear relationship is modelled using (14).

$$O_t = CL_t + \varsigma_t \tag{14}$$

where O_t is the set of observations for the leaf at time instant t, ς is Gaussian random noise and C represents a constant matrix.

The functionality of the RBPF is similar to a generic particle filter, where the posterior density is represented by a set of weighted particles, $S_t = \{s_t^i, \omega_t^i | 1 \leq i \leq N\}$. Each represents a particle from $p(R_t|Z_t)$ and $p(L_t|R_t^i, Z_t)$. Therefore each particle is represented by $s_t^i =< R_t^i, \mu_t^i, \sigma_t^i >$. The RBPF algorithm will approximate the non-linear component involving the root variables using a particle filter, while apply Kalman filter to estimate the scale parameters which are conditional on the root variables. The proposed enhancements in the RBPF framework is presented through detailed description of the individual models in the following sections.

3.2.1. Motion Model

In general, particles are propagated at each time step using (11). The target motion is modelled using the Bayesian expansion given in (15).

$$p(R_t|R_t^i, Z_t) = p(Z_t|R_t)p(R_t|R_t^i) \tag{15}$$

After this step, an a-priori estimate of the variables $< \cap R_t^i, *, * >$ is produced. In conventional particle filtering, the observation model is applied directly after this step to estimate likelihood. However, in the RBPF framework, a Kalman prediction for the leaf variables is initiated according to (16).

$$p(L^t | T_m^t, R_m^{t-1}, L_m^{t-1}, Z^t) \tag{16}$$

The state of the Kalman filter is propagated using (12). The estimates of the Kalman filter is performed in a manner similar to [26] and is described in (17),

$$\begin{aligned} \hat{H}_t^i &= H_{t-1}^i + \beta(\hat{y}_t^i - y_{1-1}^i)/\gamma \\ \hat{W}_t^i &= \frac{\hat{H}_t^i}{H_{t-1}^i} W_{t-1}^i \\ \hat{\sigma}_t^i &= A\sigma_{t-1}^i A' + P \\ \hat{O}_t^i &= C\hat{\mu}_t^i \end{aligned} \tag{17}$$

where the parameters β and γ control the scale change rate of the ellipse with respect to the motion of the target.

3.2.2. Observation Model

The next step in the E-RBPF algorithm is the application of the observation model to measure likelihood. One main novelty of our E-RBPF framework is the tight integration of results of the proposed detection technique and tracking likelihood to model the observations. The concept of the observation model is to evaluate for each particle is given in (18).

$$\omega_t^i = p(Z_t | R_t^i, \mu_t^i, Z_{1:t}) \tag{18}$$

In order to compute the weights (ω_t^i), three main component information are combined: a) tracker confidence, b) detector confidence and c) tracker-detector deviation. The representation of the combination is as follows (19),

$$\varpi_t^i = \varpi_1 . G_h^* . G_g^* + \varpi_2 . p(\Re - \mho) + \varpi_3 . \vartheta . p(\Im - \Re) \tag{19}$$

where $(\varpi_1, \varpi_2, \varpi_3)$ are constants, component $G_h^*.G_g^*$ represents the tracker confidence, $p(\Re - \mho)$ is detector confidence, and $p(\Im - \Re)$ represents tracker-detector deviation.

Tracker Confidence: The tracker confidence term estimates the likelihood of each particle using color histogram and gradient features. This component of the weights is independent of the other weights as used in conventional tracking algorithms.

a) Similarity between color histograms of the particle and the target regions is estimated using (20),

$$G_h^* = \frac{1}{\sqrt{2\pi}\sigma_c} \exp(-\frac{1-\rho[s_\Gamma^i, r]^2}{2\sigma_c^2}) \tag{20}$$

where $\rho[\Gamma^i, r]$ measures the similarity (Kullback-Leibler Divergence) between the color histogram Γ for each particle i (ellipse) characterised by $(x_t^i, y_t^i, H_t^i, W_t^i)$ using (21),

$$\Gamma^i = f \sum k\{\frac{||(x_t^i, y_t^i) - \theta^i||}{a}\}\delta^*[h(\theta^i) - u] \tag{21}$$

where δ^* is the Kronecker delta function and $h(\theta_n)$ is a bin-assignment function at each location characterised by θ_n and color u. k weights the pixels closer to the center higher than others.

b) The second sub-component of the tracker confidence consists of measuring the gradient difference between the particles and the target. The weight from the gradient component is considered as given in (22).

$$G_g^* = \frac{1}{\sqrt{2\pi}\sigma_g} \exp(-\frac{1-g(\Gamma^i)^2}{2\sigma_g^2}) \tag{22}$$

Detector Confidence: In order to compute the detector confidence, a trained model based method as in [27] is evaluated on the image patch defined at the location of the particle with the corresponding size. Such an evaluation produces a detector output referred to as $\mho$. Further, the (dis)similarity between the

detection using the proposed algorithm $\Re$ and the model-based detector output $\mho$ is estimated as the detector confidence $p(\Re - \mho)$. This confidence term uses color and gradient information as aforementioned to assess the detection process against a model-specific method.

Tracker-Detector Deviation: The final factor evaluates the (dis)agreement between the output of the detector and the tracker at the location specified by the particle with the corresponding size. This computes the distance between the particle prediction of the target denoted as $\Im$, and the detection ($\Re$) using the proposed algorithm. Such a factor enables the robust guiding of particles. This is attained by associating one detection to each target and implemented using a matching score metric for each pair $(\Im, \Re*)$. The maximum score is iteratively selected and only the associated detections with matching score higher than a pre-defined threshold are used for association. A detailed description of the matching score matrix is presented in 3.2.6.

3.2.3. Re-Sampling

Re-sampling of particles is performed to create a new particle set such that mismatches can be corrected and also to avoid degeneracy of particles. One common approach is through replacement by weighing particles and re-sampling according to those weights. The resulting particle set indeed approximates the target distribution. Such a re-sampling technique can be represented using (23).

$$p(< R_t^i, \mu_t^{i-}, \sigma_t^{i-} >=< R_t^{j-}, \mu_t^{j-}, \sigma_t^{j-} >) \propto \varpi_t^j \tag{23}$$

The prediction PDF is modified using the newest measurements for the root as well as leaf variables. Following this step, the new samples are of the form $s_t^i =< R_t^i, \mu_t^{i-}, \sigma_t^{i-} >$.

The re-sampling mechanism proposed in this paper is based on the replacement of particles with low weights with particles regenerated based on the location of targets as detected by the detector nearest to the current location of the target as being tracked.

3.2.4. Update

Kalman update is accomplished using (24),

$$\begin{aligned} K_t^i &= \sigma_t^{i-} C^{'} (C\sigma_t^{i-} C^{'} + Q)^{-1} \\ \mu_t^i &= \mu_t^{i-} + K_t^i (O_{t-1} - C\mu_t^{i-}) \\ \sigma_t^i &= \sigma_t^{i-} - K_t^i C\sigma_t^{i-} \end{aligned} \tag{24}$$

where K_t^i is the Kalman gain that aims to minimize the posterior error covariance. Here, the samples are updated into the form $s_t^i =< R_t^i, \mu_t^i, \sigma_t^i >$. Further, the mean state is computed by averaging the state particles in the manner mentioned in (25).

$$E[S_t] = \frac{\sum_{i=1}^{N} s_t^i}{N} \tag{25}$$

3.2.5. Noise Model

The choice of the noise model can play a crucial role in the accurate localization and tracking of targets, particularly in the presence of clutter. This paper proposes the distinguished choice of noise models for the location and appearance components of the target. That is, Gaussian noise is modelled for both location and appearance; assuming independence between the two. Mathematically, it can be represented using (26).

$$p(\mathbf{x}|\mathbf{z}, \oslash_L, \oslash_A) = p(\mathbf{x}_L|\mathbf{z}_L, \oslash_L).p(\mathbf{x}_A|\mathbf{z}_A, \oslash_A) \tag{26}$$

A Gaussian noise with zero mean and scalar covariance is considered for both the location and appearance parameters of the target in the form given in (27),

$$\begin{aligned} p(\mathbf{x}_L|\mathbf{z}_L) \ N(0, \Sigma_L = \sigma_L.I) \\ p(\mathbf{x}_A|\mathbf{z}_A) \ N(0, \Sigma_A = \sigma_A.I) \end{aligned} \tag{27}$$

using parameters $\oslash = (\sigma_L, \sigma_A)$. Therefore the conditional probability can be reduced to (28).

$$p(\mathbf{x}|\mathbf{z}, \oslash) = \frac{1}{2\pi\sigma_L^2} e^{-\frac{||\mathbf{x}_L - \mathbf{z}_L||^2}{2\sigma_L^2}} \cdot \frac{1}{2\pi\sigma_A^2} e^{-\frac{||\mathbf{x}_A - \mathbf{z}_A||^2}{2\sigma_A^2}} \tag{28}$$

This produces two main parameters σ_L and σ_A that require estimation. Due to the independence assumed between location and appearance, each of these parameters can be separately estimated using a Maximum Likelihood formulation. Parameter estimation is carried out such that the convergence guarantees an increase of the likelihood.

3.2.6. Data Association

With the presence of noisy measurements and multiple targets, it is important to associate one detection to at most one target, and thereby solve the data association problem. The association algorithm computes a matching score matrix for each pair (τ^*, ϱ^*), where τ^* represents all tracked targets and ϱ^* refers to detector outputs. The matching function estimates the distance between particles of tracked target (T) at various detections (ϱ^*) as given in (29),

$$\hbar(\Im^*, \Re^*) = \aleph(\Im^*, \Re^*).p(\Im - \Re^*) \tag{29}$$

where $\aleph(\Im^*, \Re^*)$ is the gating function and $p(\Im - \Re^*)$ is the same component that measures the tracker-detector deviation mentioned earlier. Note that, the tracker-detector deviation is measured for each tracked target only once and used for both the likelihood computation and data association.

Gating Function: In addition to the distance between the detection and tracker, the gating function assesses each detection based on its location with respect to the velocity and direction of the target. The probabilistic gating function can be represented using (30).

$$\aleph(\tau^*, \varrho^*) = p(size|\Im^*)p(pos|\Im^*) \tag{30}$$

4. Experiments & Analysis

In this section of the paper, experimental details evaluating the proposed model and comparing it against competing baseline techniques are presented. Two types of experiments are conducted: 1) that evaluates the detection and tracking in a mutually exclusive manner and 2) in an integrated fashion. Further, all the methods are validated on a wide range of both real and synthetic datasets using standard performance metrics. For experimental evaluation, 12 real scenes and 24 synthetic sequences have been chosen from a variety of publicly available datasets. The real scenes include sequences from the PETS 2001, PETS 2004, and PETS 2006 datasets [28] captured both in indoor and outdoor environments. The scenario depicted in these sequences demonstrate dynamic illumination that have been categorized into 5 distinct levels of difficulty: a) normal (N) consisting of no dominant illumination changes, b) difficulty level 1 (D1) including gradual changes in the illumination, either increasing or decreasing, c) difficulty level 2 (D2) involving abrupt changes in the illumination introduced in a periodic manner, d) difficulty level 3 (D3) with a combination of short, gradual, and abrupt changes in the illumination introduced in an irregular manner and e) difficulty level 4 (D4) containing randomly introduced illumination changes against dynamic backgrounds. From a tracking perspective, these videos are chosen to contain moving targets, mostly people and vehicles, encompassing a range of tracking complexities including occlusion, camera shake/jitter, false alarms, motion dynamics, etc. All datasets contain ground truth annotation for both target detection and tracking.

Experimental validation is performed through both on qualitative and quantitative evaluations. For qualitative evaluation, results of detection and tracking are presented as output frames to be verified through visual inspection. On the other hand, in order to quantitatively evaluate and benchmark our detection technique, the precision-recall ratio, F-measure and PSNR metrics are used.

These measures are described in Equations (31-34),

$$Precision = \frac{TP}{TP + FP} \tag{31}$$

$$Recall = \frac{TP}{TP + FN} \tag{32}$$

$$F - Measure = 2.\frac{Precision.Recall}{Precision + Recall} \tag{33}$$

$$PSNR = 10.\log_{10}(\frac{MAX_I^2}{MSE}) \tag{34}$$

where TP (true positives), FP (false positives) and FN (false negatives) denote those number of pixels correctly detected, detected incorrectly, and undetected but incorrectly, respectively. The F-measure is a metric to determine the accuracy of the tests that is derived quantitatively as a balance between precision and recall, wherein, a higher value indicates better accuracy. PSNR is measured as the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Here, MAX_I is the maximum possible pixel value of the image and MSE denotes the Mean Squared Error.

In order to benchmark the tracking procedure through quantitative evaluation, the use of two popular performance metrics from the tracking domain [29] namely: Multiple object tracking precision (MOTP) and b) Multiple object tracking accuracy (MOTA) is proposed. These estimates are based on determining the distance d_t^τ between the target τ and its corresponding hypothesis at each instant of time t. The MOTP and MOTA measures are described in Equations (35-36),

$$MOTP = \frac{\sum_{t,\tau} d_t^\tau}{\sum_t c_t^\tau} \tag{35}$$

$$MOTA = 1 - \frac{\sum_t (m_t^\tau + f_t^\tau + h_t^\tau)}{\sum_t q_t^\tau} \tag{36}$$

where $c_t\tau$ represents the number of matches found at each instant of time t, m_t^τ,f_t^τ and h_t^τ represent the misses, false positives and mismatches, respectively, and q_t^τ represents the number of targets present at time instant t.

In addition to these measurements, the following metrics inspired by the work of [30] are also computed. a) Mostly Tracked (MT)%: Percentage of the ground truth trajectories which are covered by the tracker output for more than 80% in length, b) Mostly Lost (ML)%: Percentage of ground truth trajectories which are covered by the tracker output for less than 20% in length (smaller the better), c) Fragments (Fr): The total number of times that a ground truth trajectory is interrupted by the tracker (smaller the better), d) ID Switches (IDS): The total number of times that a tracker trajectory changes its matched ground truth identity (smaller the better) and f) Root Mean Squared Error (RMSE) error: The difference between the target and ground truth trajectories on the (MT) trajectories.

A selection of 3 baseline methods are chosen to benchmark the proposed DRA-base hybrid target detection framework. These include: the famous adaptive background modelling of [31], a spatially adaptive illumination modelling technique for background subtraction proposed by [32] and the background modelling technique based on bi-directional analysis as in [33]. The proposed method can be distinguished from these baselines in the manner as follows: a) the proposed technique is based on the reverse analysis of frames for building robust background models as against analysing pixel changes only in forward direction as shown in [31], [32] b) the proposed technique does not only perform reverse analysis to pick the best of the two (forward and backward) results as in [33], however, builds a hybrid model using a selected composition of frames from the forward and reverse directions using sufficient statistics, c) in addition to

autonomously detecting the number of frames required from the forward and reverse directions to build the hybrid model, the proposed method also provides a mechanism of self-adaptation through implicitly learning the changes in illumination conditions that is not feasible with the baseline techniques proposed in [31], [32] and [33] and d) the proposed method also makes fewer assumptions on the nature of illumination changes and hence is more generic to the changes in the real-world scenarios.

The proposed E-RBPF model is compared to 4 state-of-the-art tracking techniques including: a) Generic RBPF tracker (G-RBPF) for multiple target tracking [34], b) a Probabilistic Data Association particle Filtering (PDAF) technique for multiple object tracking proposed in [35], c) an extended version of the context tracking (CT) algorithm proposed in [36] and c) the locally orderless tracking (LOT) from [37]. The relevance of these baseline trackers to the proposed E-RBPF model can be described as follows. In comparison to the G-RBRF and PDAF trackers, the usefulness of integrating the extended detect-track likelihood model, data association and noise models into a unified joint E-RBPF framework are demonstrated. Further, the differences in the use of similar noise models in comparison against the Locally Orderless Tracker (LOT) are also demonstrated.

4.1. Detection Results

In Figure 5, the results of target detection on one frame of a D3 category sequence (in row 1) and two frames of a D4 category sequence (in row 2 and row 3) are compared. As it can be observed, the results of the proposed hybrid detection technique in Figure 5(c), Figure 5(g) and Figure 5(k) resemble the ground truth in Figure 5(b), Figure 5(f) and 5(j) more closely than the adaptive background modelling of [31] counterpart in Figure 5(d), Figure 5(h) and Figure 5(l). The superiority in the performance of the proposed method can be attributed to the selective composition of frames from the forward and reverse directions that has facilitated building a more robust and accurate background model in

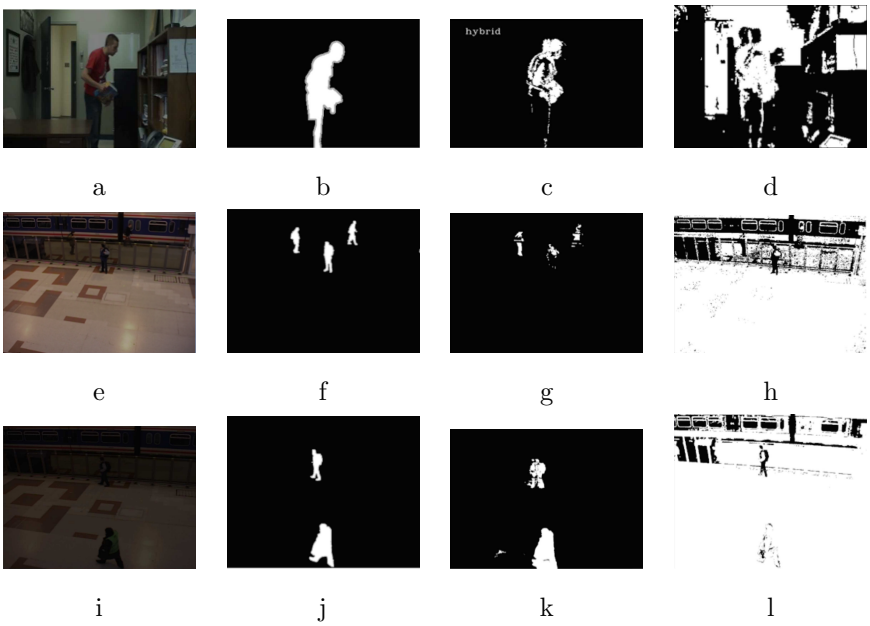


Figure 5: Comparative results of the proposed DRA-based hybrid background modelling method (dubbed as hybrid) (c&g&k) against the adaptive background modelling method of [31] (d&h&l) and the corresponding ground truth (b&f&j) on one original frame (a) of a D3 category sequence and two original frames (e&i) of a D4 category sequence.

comparison to the baseline method of [31] that has demonstrated poorer segmentation with respect to the ground truth and the proposed hybrid method.

For quantitative evaluation, the proposed hybrid detection method is benchmarked against baselines on different frames using Recall and Precision in Table 1, and using F-measure and PSNR in Table 2 on selected frames of the same D4 category sequence as used in Figure 5. The comparative evaluation on recall and precision from Table 1 shows consistent superiority of the proposed algorithm against competing baselines on a majority of the frames. A similar trend in performance can also be noted using the F-measure and the PSNR, wherein, the proposed method outperforms its competing baselines in Table 2.

Furthermore, the proposed framework was also tested on the collective dataset

Frame	Recall				Precision			
	Hybrid	[31]	[33]	[32]	Hybrid	[31]	[33]	[32]
50	**0.632**	0.595	0.615	0.598	**0.579**	0.513	0.551	0.565
75	**0.767**	0.760	0.762	0.760	**0.635**	0.556	0.603	0.598
100	**0.720**	0.644	0.707	0.679	**0.730**	0.519	0.716	0.676
125	**0.685**	0.671	0.680	0.683	**0.670**	0.548	0.665	0.587

Table 1: Comparison of quantitative performance of the proposed detection technique (dubbed as Hybrid) against state-of-the-art methods ([31], [33] & [32]) using the Precision and Recall metrics on selected frames (50,75,100,125) of the D4 category sequence used in Figure 5.

Frame	F-Measure				PSNR			
	Hybrid	[31]	[33]	[32]	Hybrid	[31]	[33]	[32]
50	**0.604**	0.551	0.581	0.581	**30.60**	16.17	27.65	26.51
75	**0.695**	0.643	0.673	0.669	**31.49**	22.40	29.63	25.38
100	**0.725**	0.575	0.711	0.677	**37.10**	15.66	33.12	30.76
125	**0.677**	0.603	0.603	0.631	**35.16**	24.40	34.06	31.82

Table 2: Comparison of quantitative performance of the proposed detection technique (dubbed as Hybrid) against state-of-the-art methods ([31], [33] & [32]) using the F-measure and PSNR metrics on selected frames (50,75,100,125) of the D4 category sequence used in Figure 5.

consisting of sequences at various levels of difficulty. The comparative results of the proposed and baseline techniques using the Recall and Precision metrics are presented in Table 3 and using the F-measure and PSNR are presented in Table 4. As one would anticipate, the performance of all the models deteriorate with increasing difficulty. However, in comparison to the state-of-the-art methods, the proposed technique is capable of producing more accurate and robust detection particularly with the D2, D3 and D4 category sequences.

Level	Recall				Precision			
	Hybrid	[31]	[33]	[32]	Hybrid	[31]	[33]	[32]
N	0.752	**0.755**	0.754	0.753	0.876	**0.878**	0.875	0.876
D1	0.646	**0.649**	0.647	0.647	0.857	**0.866**	0.860	**0.866**
D2	**0.610**	0.594	0.598	0.603	**0.619**	0.533	0.582	0.567
D3	**0.597**	0.576	0.583	0.578	**0.593**	0.520	0.551	0.548
D4	**0.603**	0.576	0.579	0.591	**0.568**	0.519	0.536	0.534

Table 3: Comparison of quantitative performance of the proposed detection technique (dubbed as Hybrid) against state-of-the-art methods ([31], [33] & [32]) using the Precision and Recall metrics averaged on all frames of the various categories of sequences with increasing complexity from N to D4.

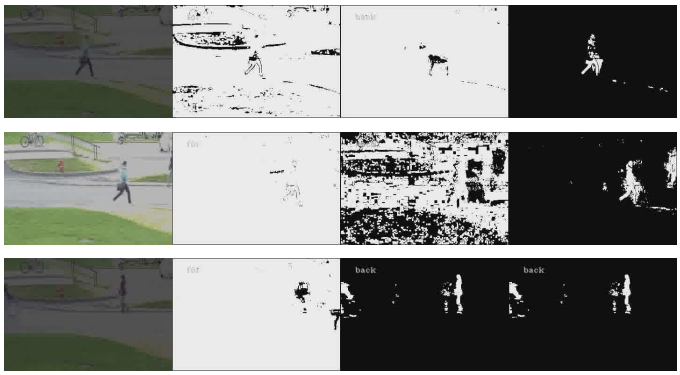
Level	F-Measure				PSNR			
	Hybrid	[31]	[33]	[32]	Hybrid	[31]	[33]	[32]
N	0.809	**0.812**	0.810	0.810	29.44	**29.50**	29.00	29.26
D1	0.737	**0.742**	0.738	0.740	28.07	28.22	**28.30**	28.24
D2	**0.615**	0.562	0.590	0.584	**23.02**	13.54	20.62	19.85
D3	**0.595**	0.547	0.566	0.562	**21.92**	9.07	18.16	12.56
D4	**0.585**	0.547	0.557	0.561	**19.50**	8.14	12.58	10.75

Table 4: Comparison of quantitative performance of the proposed detection technique (dubbed as Hybrid) against state-of-the-art methods ([31], [33] & [32]) using the F-measure and PSNR metrics averaged on all frames of the various categories of sequences with increasing complexity from N to D4.

Difficulty	Forward	Backward	Hybrid
N	91%	8%	1%
D1	76%	12%	12%
D2	65%	16%	19%
D3	23%	28%	49%
D4	13%	16%	71%

Table 5: Comparison of the percentage number of time that the forward, background and hybrid models are being built by the proposed detection technique against increasing complexity of video sequences through from N to D4.

As indicated earlier, one key novelty of the proposed detection technique is in building a hybrid background model using a selected composition of frames from the forward and backward directions. During experimentation, as observed in Table 3 and Table 4, with increasing complexity in illumination conditions, the hybrid background model is built more frequently than the models in either the forward or backward directions. In Table 5, the percentage frequency of each type of background model being built against increasing complexity of video sequences is presented. It can be proven beyond doubt that the role of hybrid background model becomes apparent with increasing complexity of the sequences. In particular, for the D4 category video sequence, in order to guarantee better detection results, frames from both forward and reverse directions are hybridized at times nearly 50% more often than in either directions individually. The improved target detection as depicted in Figure 6, solicits all previous claims of the hybrid model being more effective in representing the dynamic background.



a b c d

Figure 6: Results of the proposed target detection technique demonstrating the capacity of the hybrid model (column d) against forward (column b) and backward (column c) directional models on selected original frames (357,463&647) (column a) of a D4 category sequence.

4.2. Tracking Results

In this section, results comparing the proposed E-RBPF tracker against baseline trackers are presented. In Figure 7, the qualitative tracking results comparing the E-RBPF framework against the baseline methods: G-RBPF, PDAF, CT and LOT are illustrated. The results in Figure 7 shows tracker outputs on selected frames (represented as columns) of video sequences with increasing levels of difficulty through from D2 to D4 (across different rows). The results in Figure 7 clearly demonstrate the superiority of the proposed E-RBPF tracker against other baselines. The CT and LOT trackers produce more comparable results to the proposed method as against the $G-RBPF$ and the $PDAF$ trackers. With increasing levels of difficulty, the baseline methods, particularly the $G-RBPF$ and the $PDAF$ trackers are more susceptible to drift than the other trackers. The E-RBPF framework has proven to remain accurate in target localization despite abrupt changes to illumination conditions, occlusion and clutter.

Additionally, the qualitative results of the proposed E-RBPF method (red bound-

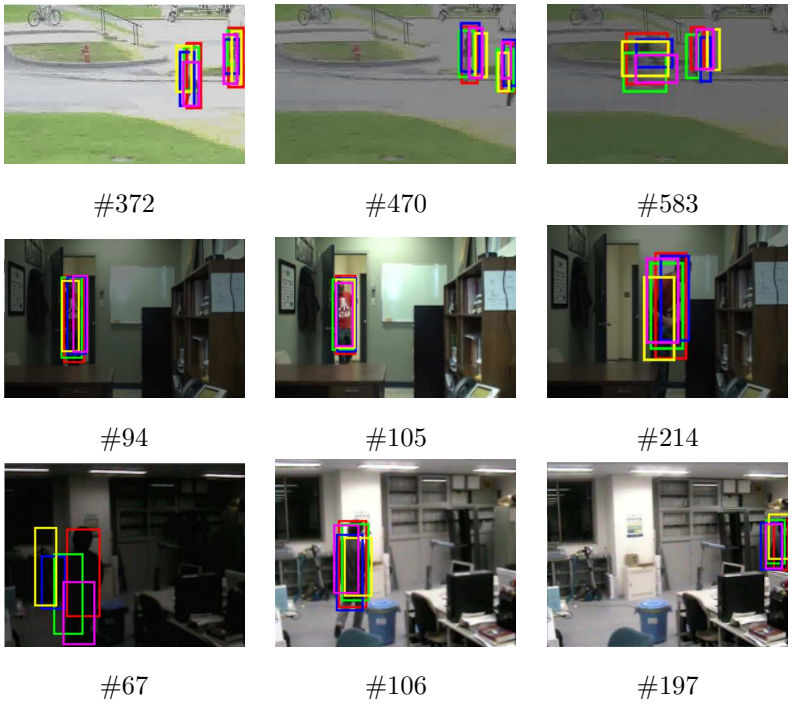


Figure 7: Qualitative target tracking results of the E-RBPF method (red bounding box) compared against the baseline methods: G-RBPF (yellow bounding box), PDAF (magenta bounding box), CT (blue bounding box) and LOT (green bounding box) on multiple frame samples (columns) on video sequences of increasing levels of difficulty through from D2 (row 1), D3 (row 2) and D4 (row 3).

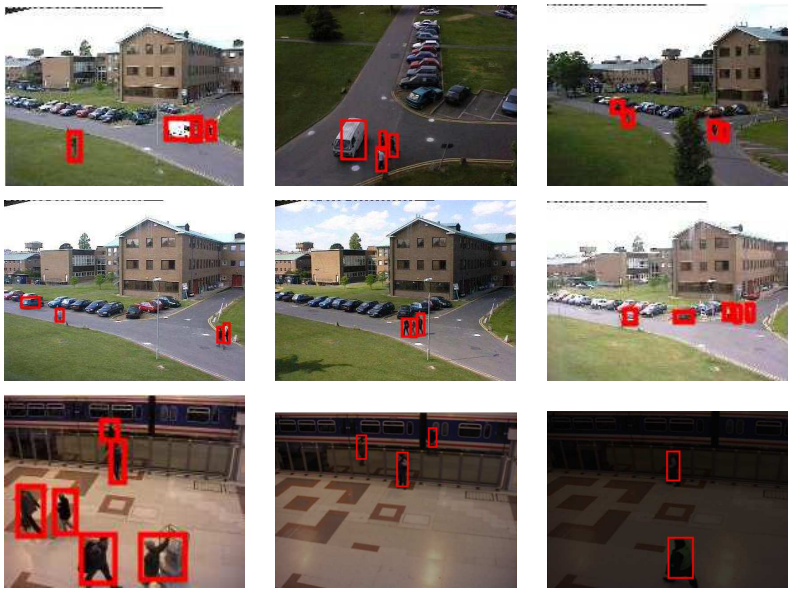


Figure 8: Target tracking results of the E-RBPF method (red bounding box) on multiple frame samples from the PETS database [28].

ing box) on example frames from video sequences consisting of multiple moving targets (both humans and vehicles) captured at various perspectives and scales, under varying real-world illumination conditions from PETS 2001, PETS 2004, and PETS 2006 sequences can be found in Figure 8. A quick visual inspection of these results reveals the superior accuracy and robustness of the proposed E-RBPF tracker.

In addition to qualitative comparison, the trackers are benchmarked against each other using quantitative evaluation metrics which are summarized in Table 6.

It is important to note that the success of the trackers as outline above can be mainly attributed to the accurate localization of targets during the detection phase. In order to quantify the influence of the hybrid detection strategy on the tracking results, competitive tracking results are generated by initializing

Method	MOTA	MOTP	MT	ML	IDS	FR
ERBPF	**73.2%**	**84.98%**	**53.73%**	**27.61%**	**9**	**312**
G-RBPF	55.61%	73.52%	33.41%	29.43%	23	427
PDAF	56.19%	71.56%	31.53%	33.69%	27	468
CT	61.58%	77.88%	42.12%	32.06%	18	394
LOT	60.36%	79.91%	47.92%	35.85%	19	409

Table 6: Quantitative evaluation comparing the proposed E-RBPF tracker against the baseline trackers (all initialized using the proposed hybrid detector) using various performance metrics averaged on all frames of the different categories of video sequences.

Method	MOTA	MOTP	MT	ML	IDS	FR
ERBPF	**59.87%**	**66.34%**	**37.62%**	**57.61%**	**15**	**402**
G-RBPF	36.12%	41.79%	17.93%	66.87%	27	519
PDAF	39.83%	44.27%	18.18%	69.98%	36	527
CT	41.69%	51.38%	27.51%	66.17%	23	486
LOT	43.55%	54.62%	24.47%	71.28%	26	491

Table 7: Quantitative evaluation comparing the proposed E-RBPF tracker against the baseline trackers (all initialized using the baseline detection technique of [33]) using various performance metrics averaged on all frames of the different categories of video sequences.

the tracker with the baseline detection method of [33]. The differences in the results of tracking are presented in Table 7 and is itself indicative of the power of the proposed detection framework.

In Table 8, the RMSE error between the tracked targets and the ground truth MT trajectories are detailed. This measure is indicative of the deviation of the target trajectory from the ground truth.

Finally, with regards to the comparison of computational demand of the proposed hybrid detection strategy against its uni-directional counterpart, intu-

Method	D1	D2	D3	D4
ERBPF	**38**	**42**	**17**	**36**
G-RBPF	128	143	109	227
PDAF	134	150	127	186
CT	112	138	126	107
LOT	64	97	73	59

Table 8: Comparison of RMSE error of the predicted target trajectory estimated using all trackers (initialized using the proposed hybrid detector) against its corresponding MT ground truth trajectory averaged across all video sequences from each categorized levels of difficulty D1 to D4 (columns).

itively nearly 3 times overhead is expected. However, with an optimized circular buffer implementation for processing the learning frames, it has been possible to reduce the computational overhead to approximately 1.5 times existing uni-directional background modelling schemes. At this point, a judicious decision on the trade-off between the accuracy of detection to the computational demands requires to be made. In the context of the detector having to initialize the tracking algorithm (as proposed), little such compromise can be made on the accurate detection of the targets. However, a more relaxed detection may suffice the needs of the likelihood measurement step during tracking. During tracking, comparable computational requirements between the proposed E-RBPF and the G-RBPF could be observed assuming a-priori detection. Tests indicate that the the tracking procedure in its unoptimized MATLAB implementation with 1000 particles and 100 iterative cycles can converge to 2 fps of tracked output in real-time.

5. Conclusion

In this paper, a method that seamlessly integrates a DRA-based background modelling mechanism for target detection with a E-RBPF tracking framework for accurate target localization under the presence of illumination changes, occlusion and camera shake is proposed. The results of comparing the proposed model against baselines has shown significant improvements both quantitatively and qualitatively. The future of this research is to extend the model for tracking large number of targets in crowded scenes and across multiple cameras.

References

[1] M. Han, A.Sethi, W. Hua, Y. Gong, A detection-based multiple object tracking method, in: In the Proc. of International Conference on Image Processing (ICIP), Vol. 5, 2004, pp. 3065–3068.

[2] S.J.Davey, M.G.Rutten, B.Cheung, A comparison of detection performance

for several track-before-detect algorithms, in: In the Proc. of International Conference on Information Fusion (ICIF), 2008, pp. 1–8.

[3] Y.Boers, H.Driessen, J.Torstensson, M.Trieb, R.Karlsson, F.Gustafsson, Track-before-detect algorithm for tracking extended targets, In the Proc. of IEE Radar, Sonar and Navigation 153 (4) (2006) 345–351.

[4] J. Wu, S. Hu, Y. Wang, Adaptive multifeature visual tracking in a probability-hypothesis-density filtering framework, Signal Processing 93 (11) (2013) 2915–2926.

[5] R. E. Bethel, B. Shapo, C. M. Kreucher, {PDF} target detection and tracking, Signal Processing 90 (7) (2010) 2164 – 2176.

[6] Y. Li, H. Xiao, Z. Song, R. Hu, H. Fan, A new multiple extended target tracking algorithm using {PHD} filter, Signal Processing 93 (12) (2013) 3578–3588, special Issue on Advances in Sensor Array Processing in Memory of Alex B. Gershman.

[7] X. Zhou, X. Li, Dynamic spatio-temporal modeling for example-based human silhouette recovery, Signal Processing 110 (0) (2015) 27–36.

[8] A. Yilmaz, O. Javed, M. Shah, Object tracking: A survey, ACM Computing Surveys 38 (4).

[9] S. Das, N.Vaswani, Particle filtered modified compressive sensing (pafimocs) for tracking signal sequences, in: In the Proc. of Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2010, pp. 354–358.

[10] G. Yu, H. Lu, Illumination invariant object tracking with incremental subspace learning, in: In the Proc. of International Conference on Image and Graphics (ICIG), 2009, pp. 131–136.

[11] P. Prez, C. Hue, J. Vermaak, M. Gangnet, Color-based probabilistic tracking, in: In the Proc. of European Conference on Computer Vision (ECCV), 2002, pp. 661–675.

[12] F.Moreno-Noguer, A. Sanfeliu, D.Samaras, A target dependent colorspace for robust tracking, in: In the Proc. of International Conference on Pattern Recognition (ICPR), Vol. 3, 2006, pp. 43–46.

[13] M. Cristani, M. Farenzena, D. Bloisi, V. Murino, Background subtraction for automated multisensor surveillance: A comprehensive review, EURASIP Journal of Advanced Signal Processing 2010 (2010) 43:1–43:24.

[14] T. Bouwmans, Traditional and recent approaches in background modeling for foreground detection: An overview, Computer Science Review II (I2).

[15] H. Veeraraghavan, P. Schrater, N. Papanikolopoulos, Robust target detection and tracking through integration of motion, color, and geometry, Computer Vision and Image Understanding 103 (2) (2006) 121 – 138.

[16] G. Silveira, E.Malis, Real-time visual tracking under arbitrary illumination changes, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–6.

[17] Z. Khan, T. Balch, F. Dellaert, A rao-blackwellized particle filter for eigentracking, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2004, pp. 980–986.

[18] X. Xu, B. Li, Rao-blackwellised particle filter for tracking with application in visual surveillance, in: In the Proc. of IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 17–24.

[19] J. Pan, B. Hu, J. Q. Zhang, Robust and accurate object tracking under various types of occlusions, IEEE Transactions on Circuits and Systems for Video Technology 18 (2) (2008) 223–236.

[20] P. Chavali, A. Nehorai, Hierarchical particle filtering for multi-modal data fusion with application to multiple-target tracking, Signal Processing 97 (0) (2014) 207–220.

[21] H.Bhaskar, L.Mihaylova, A.Achim, Video foreground detection based on symmetric alpha-stable mixture models, IEEE Transactions on Circuits and Systems for Video Technology 20 (8) (2010) 1133–1138.

[22] J.Batista, P.Peixoto, C.Fernandes, M.Ribeiro, A dual-stage robust vehicle detection and tracking for real-time traffic monitoring, in: In the Proc. of Intelligent Transportation Systems Conference (ITSC), 2006, pp. 528–535.

[23] A. Romanoni, M. Matteucci, D. G.Sorrenti, Background subtraction by combining temporal and spatio-temporal histograms in the presence of camera movement, Machine Vision and Applications 25 (6) (2014) 1573–1584.

[24] A.Mittal, N.Paragios, Motion-based background subtraction using adaptive kernel density estimation, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2004, pp. 302–309.

[25] G. Kovai, L. Tao, D. Cai, Michael.J.Shelley, Theoretical analysis of reverse-time correlation for idealized orientation tuning dynamics, Journal of Computational Neuroscience 25 (3) (2008) 401–438.

[26] X. Xu, B. Li, Adaptive raoblackwellized particle filter and its evaluation for tracking in surveillance, IEEE Transactions on Image Processing 16 (3) (2007) 838–849.

[27] F.Han, Y.Shan, R.Cekander, H.Sawhney, R.Kumar, A two-stage approach to people and vehicle detection with hog-based svm, in: In the Proc. of IEEE International Workshop on Performance Metrics for Intelligent Systems, 2006.

[28] Dataset - pets: Performance evaluation of tracking and surveillance (2000-2014).
URL `http://www.cvg.rdg.ac.uk/slides/pets.html`

[29] K. Bernardin, R. Stiefelhagen, Evaluating multiple object tracking performance:the clear mot metrics, EURASIP Journal on Image and Video Processing 2008.

[30] Y. Li, C. Huang, R. Nevatia, Learning to associate: Hybridboosted multi-target tracker for crowded scene, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2953–2960.

[31] C. Stauffer, W. E. L. Grimson, Adaptive Background Mixture Models for Real-Time Tracking, in: Computer Vision and Pattern Recognition, Vol. 2, 1999, pp. 2246–2252.

[32] J.K.Paruchuri, E.P.Sathiyamoorthy, S.S.Cheung, C.-H. Chen, Spatially adaptive illumination modeling for background subtraction, in: In the Proc. of IEEE International Conference on Computer Vision Workshops (ICCV), 2011, pp. 1745–1752.

[33] A.Shimada, H.Nagahara, R.I.Taniguchi, Background modeling based on bidirectional analysis, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1979–1986.

[34] S. Srkk, A. Vehtari, J. Lampinen, Rao-blackwellized particle filter for multiple target tracking, Information Fusion Journal 8 (2005) 2007.

[35] M.Ekman, Particle filters and data association for multi-target tracking, in: In the Proc. of International Conference on Information Fusion (ICIF), 2008, pp. 1–8.

[36] T. B. Dinh, N. Vo, G. Medioni, Context tracker: Exploring supporters and distracters in unconstrained environments, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1177–1184.

[37] S. Oron, Locally orderless tracking, in: In the Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1940–1947.