Tracking Ground Based Targets in Aerial Video with Dual-Tree Wavelet Polar Matching and Particle Filtering

J. D. B. Nelson, S. K. Pang, N. G. Kingsbury, and S. J. Godsill
University of Cambridge
Signal Processing and Communications Laboratory
CUED, Trumpington Street, Cambridge CB2 1PZ
{jdbn2, skp31, ngk, sjg}@eng.cam.ac.uk

***Abstract*—The detection and tracking of targets in aerial imagery of cluttered urban environments is addressed. Polar matching, using dual-tree complex wavelet transforms, is used as a shift and rotation invariant detector. A particle filter is employed to add robustness, especially in the event of target occlusion. We show that, together, these methods can robustly track a ground based target as it becomes both partially and fully occluded.**

Keywords: Tracking, Dual-Tree Complex Wavelets, Polar Matching, Particle Filtering

I. Introduction

Cluttered urban environments can present difficult and interesting problems for automatic detection and tracking of ground based targets with aerial imagery. As the location, bearing, and azimuth of the sensor change, the image of the target will appear to shift and rotate, and possibly change in scale. In this context, it therefore makes sense that any successful detection method must be invariant, or at least robust, to spatial shifts, rotations, and scale variations. To this end, the descriptor and matching technique afforded by rotation-invariant polar matching with dual-tree complex wavelet transforms (DTCWT) recently developed by Kingsbury in 2006 [8] is adapted here, to the task of detection.

Previous attempts to use wavelets to extract features have been hampered by the difficulties of incorporating such invariances. Either the target must somehow be normalised first, or the wavelet transform must be invariant to certain transformations. In practice, normalisation can be difficult. For translation, authors sometimes implement a variant of the 'spin-cycle' method of Coifman and Donoho [3] whereby extra training samples are created by shifting the original ones. A more elegant method is to construct transforms which are themselves invariant. The shiftable wavelet, introduced by Simoncelli et al [18], satisfies a slightly weaker condition than shift invariance but is less redundant than the spin cycle. Introduced by Kingsbury [7], [17], the dual-tree complex wavelet transform (DTCWT) is approximately shift invariant and offers lower redundancy with greater computational efficiency. Moreover, Kingsbury discovered a recent extension of the DTCWT, known as polar matching [8], that is also approximately rotation invariant. Unlike the DTCWT rotation invariant work of Hill et al [10], polar matching retains the phase information of the complex coefficients and therefore represents a richer descriptor. The output of the polar matching method gives a detection confidence, or likelihood value, of the target of interest for a specific position and orientation within the video frame.

In cluttered urban environments, situations can arise where objects, similar to the target, come into the field of view. More seriously, the target may even become partially or fully occluded from sensor view. For these reasons, it is a good idea to temper the descriptor with some kind of dynamic model or tracking method. Various approaches have been proposed to tackle the problem of target tracking ranging from the Kalman filter to non-linear extensions such as joint probabilistic data association filters [1] [2]. With recent developments in optimal non-linear techniques such as particle filters [6] [4] and Markov Chain Monte Carlo (MCMC) [5] [16], it is now possible to consider exploiting other information, such as non-linear measurement processes, which potentially result in significant performance gains.

We shall discuss how the detection output of the polar matching method can be used with a tracking filter to provide smooth estimates of the target's position. Optimal linear filters such as the Kalman filter may not work well in this scenario. One reason is due to the non-linear measurement process of the imaging sensor and the polar matching method. Another reason is that the posterior distribution is likely to be multimodal due to the cluttered nature of the video data. To overcome these issues, we have designed a particle filter to perform the tracking.

II. Polar matching

Extending his work on the shift-invariant dual-tree complex wavelet transform [7], Kingsbury recently introduced the rotation-invariant polar matching method [8]. Owing to low redundancy, the DTCWT descriptor is more efficient than the existing popular scale- and rotation-invariant method of SIFT [14]. It is adapted here to provide image matching between a

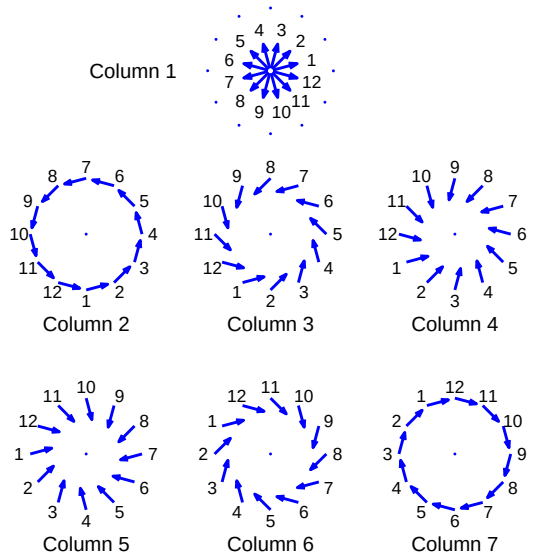

Figure 1. Locations and orientations of the DTCWT coefficients. Each orientation describes a coefficient, or conjugate, of one of the six subbands. The column numbers indicate the column of the P-matrix in which the coefficients are placed, and the numbers displayed around each circle indicate the row of the P-matrix that each coefficient is placed. Taken from [8].

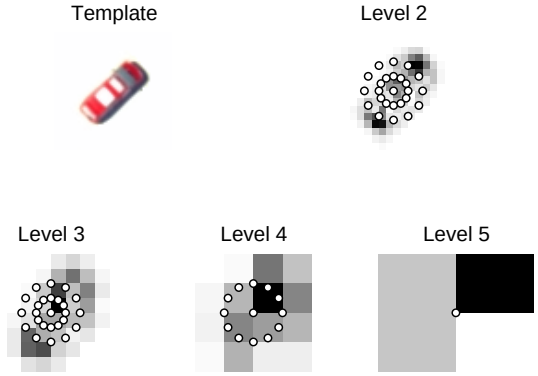

Figure 2. DTCWT coefficients of a template. The white dots indicate the locations of the coefficients that are used in the P-matrix template. Four different scales, or levels, are used. Only one of one of the 6 subbands is shown.

small template and a larger image rather than matching key-points of two similarly sized images, as previously reported.

A. Methodology

Kingsbury's method proceeds by firstly computing the DTCWT coefficients of a template. The centre of the target is located manually and the complex wavelet coefficients at this point are stored. Coefficients are also taken around one or more circles, about the centre point, at 30 degree increments and at multiple scales. As Figure 1 illustrates, the coefficients are then arranged into a polar matching matrix (P-matrix) such that a rotation of $k \times 30^\circ$ in the original image will manifest a vertical shift by k rows in the P-matrix. Consider two images, one a $30n^\circ$ rotated version of the other; then a sum of column-wise correlations between the two corresponding P-matrices will result in a response curve, with respect to relative rotation angle, and a maximum at n.

However, the rotational sensitivity can be increased to 7.5° via careful band-limited interpolation. This is achieved by performing the correlation as a product in the Fourier domain and zero padding. Care should be taken here. The first column of a P-matrix, formed about the centre of a single step edge will vary slowly as the the edge is rotated. Columns 2 and 7 will vary quicker, 3 and 6 quicker still and 4 and 5 quickest of all. Hence, the zeros must be placed according to P-matrix column. Coefficients obtained from other scales can be added by appending them as extra columns to the P-matrix. Hence, this polar matching technique takes the property of shift invariance from the DTCWT, and rotation invariance from the P-matrix construction.

Let $\mathbf{P}_A(\mathbf{r}_0) \in \mathbb{C}^{M\times N}$ be the column-wise Fourier transform of the P-matrix taken about some point $\mathbf{r}_0 = (x_0, y_0)$. Furthermore, denote the (m,n)th element by $a_{m,n}(\mathbf{r}_0)$, and the column vector $\mathbf{a} = \mathbf{a}(\mathbf{r}_0) \in \mathbb{C}^{MN}$ as $[a_{1,1}, \ldots, a_{M,1}, a_{1,2}, \ldots, a_{M,2}, \ldots a_{M,N}]^T$, and write $\mathbf{A} = \mathrm{diag}(\mathbf{a}) \in \mathbb{C}^{MN\times MN}$. Likewise, let $\mathbf{b} = \mathbf{b}(\mathbf{r}) \in \mathbb{C}^{MN}, \mathbf{B} = \mathrm{diag}(\mathbf{b})$ be the coefficients in the P-matrix of a test image taken at $\mathbf{r}$. Together, the inverse Fourier transform and zero padding can be described as a matrix operator $\mathbf{W}$. Let $\mathbf{w}_k^T$ be the kth row vector of $\mathbf{W}$ and $\mathbf{W}_k = \mathrm{diag}(\mathbf{w}_k)$. Then, the polar matching operation at the angle θ_k can be written as

$$\big(A \star B\big)(\theta_k) = \frac{\Re \mathbf{w}_k^T \mathbf{A}^* \mathbf{b}}{\|\mathbf{a}\| \, \|\mathbf{b}\|}, \tag{1}$$

where the superscript * is complex conjugation, and the real component $\Re$ is taken to return the correlation intensity. Now, consider the result of performing polar matching using a P-matrix template $\mathbf{P}_A$ and a neighbourhood of candidate p-matrices $\mathbf{P}_B(\mathbf{r})$, taken over all $\mathbf{r}$ in some region, namely

$$\big(A \star B(\mathbf{r})\big)(\theta_k) = \frac{\Re \mathbf{w}_k^T \mathbf{A}^* \mathbf{b}(\mathbf{r})}{\|\mathbf{a}\| \, \|\mathbf{b}(\mathbf{r})\|}. \tag{2}$$

The result is a correlation surface defined over $\mathbb{R}^2 \times [0, 2\pi)$ and the point

$$(\hat{\mathbf{r}}, \hat{\theta}) = \underset{(\mathbf{r},\theta_k)}{\mathrm{argmax}} \big(A \star B(\mathbf{r})\big)(\theta_k) \tag{3}$$

gives the location $\hat{\mathbf{r}}$ and orientation $\hat{\theta}$ of the closest match between the template and candidates.

B. Fusion

In the application considered here, RGB colour data is available. Consequently, a decision must be made as to when to fuse information obtained from each respective channel. The most simple low level fusion involves forming a weighted linear sum of the different channels before applying polar matching. Denoting each channel by a subscript m or n, polar matching with linear low level fusion is akin to computing

$$\mathbf{w}_k \sum_m \alpha_m \mathbf{A}_m^* \sum_n \alpha_n \mathbf{b}_n = \sum_{m,n} \alpha_m \alpha_n \left(A_m \star B_n\right). \tag{4}$$

Here, each weighted channel of the template is matched with all the weighted channels of the test image. This may not be deemed appropriate, especially if each channel are of significantly different modalities or uncorrelated. On the other hand, a basic high level fusion strategy would be to take a weighted sum of the individual channel-by-channel correlations:

$$\mathbf{w}_k \sum_m \gamma_m \mathbf{A}_m^* \mathbf{b}_m = \sum_m \gamma_m \left(A_m \star B_m \right), \tag{5}$$

so that, for example, only the red channel of the template gets correlated with the red channel of the test image. Note that this is equivalent to simply concatenating the P-matrix columns from each channel together. For this reason, we favour the high level fusion approach. The weights γ_m, are set to

$$\gamma_m = \frac{1}{\|\mathbf{a}\| \, \|\mathbf{b}(\mathbf{r})\|} \quad \forall m, \tag{6}$$

where the $\mathbf{a}$ and $\mathbf{b}$ are concatenations of the separate channels. Holding γ_m constant for all channels ensures that any differences in the influence of wavelet coefficients between channels are preserved. It is equivalent to forming one single P-matrix by appending the columns of extra channels.

III. Tracking and Reacquisition

In the interests of robustness and in the event of clutter and possible occlusion, it is necessary to combine the polar matching detector with a motion model. Two approaches will be discussed below. A simple windowed search, or maximum likelihood method is considered as a base-line method. This will be compared to the behaviour of a particle filter approach. In both methods, a template P-matrix is formed, centred on the target of interest. A search for the target in an image frame is conducted by forming P-matrices from each search point. At each point a score is then computed by correlating the columns of the P-matrix template, stored in a database, with those of the candidate, and summing the result to produce a stack of correlation surfaces, with respect to location (x, y), and orientation θ. The windowed search method proceeds by forming a window, or neighbourhood, centred about the last known position of the target. For each time frame t, correlation scores $C_t(x, y; \theta)$ are then computed at every point in this window. Finally, the maximum score with respect to (x, y) is taken as the new target location, and the maximum with respect to θ is the target orientation, relative to the original template.

The data considered here, provided via the Defence Technology Centre, is a high definition synthetic image sequence, constructed to simulate a real world military image processing scenario. Two salient challenges of the data are highlighted in Figure 3. In the first example, it can be seen that the target passes through thick, black smoke. At first, the target is partially obscured by the edge of the smoke cloud, then it becomes totally occluded for a short period of time before it gradually emerges at the other side of the smoke region. In the second example, the target becomes partially occluded as it passes by a wall and building. It is important, therefore, not only to design a method that is robust to these problems but also one that is accurate when such problems are absent.

Figure 3. The image sequence comprises at least two main challenges. The top three snapshots show the targeted red car as it passes through a thick cloud of black tyre smoke. The bottom three show the target becoming partially occluded by a building.

for t **do**
  Compute C_t over neighbourhood $N_{t-1} \bigcup \Theta_{t-1}$
  $(x_t, y_t; \theta_t) \leftarrow \operatorname{argmax} C_t(x, y, \theta)$
  if $\max C_t >$ threshold **then**
    $N_t \leftarrow$ small neighbourhood of (x_t, y_t)
    $\Theta_t \leftarrow$ small neighbourhood of θ_t
  else
    $N_t \leftarrow$ large neighbourhood of $(x_{t-1} - x_{t-2}, y_{t-1} - y_{t-2})$
    $\Theta_t \leftarrow [-\pi, \pi)$
  end if
  $t \leftarrow t + 1$
end for

Figure 4. Naive tracker

A. Baseline method

Before we discuss how particle filtering can be used to overcome such challenges, we describe the following simple heuristic in Figure 4 to address the partial and full occlusion problem. If the best correlation score is higher than the threshold parameter, the next window will be centred about the location of the best score and the set of orientation angles is restricted to an interval about the best orientation. If the score falls below the threshold value, the next window will be doubled in size and centred about the extrapolation of the previous two windows. This heuristic tool allows the target to become, temporarily, partially or fully occluded. Figure 5 shows a sequence of frames and the resulting correlation scores as the target passes into, through, and out, full occlusion. The problem here is that the choice of the threshold can be critical. If it is set too high, then the legitimate target may be ignored. The longer this happens, the more out of date that the extrapolation becomes and the more likely the target will be lost. On the other hand, if the threshold is set too low, then, when the target becomes occluded, other nearby objects will

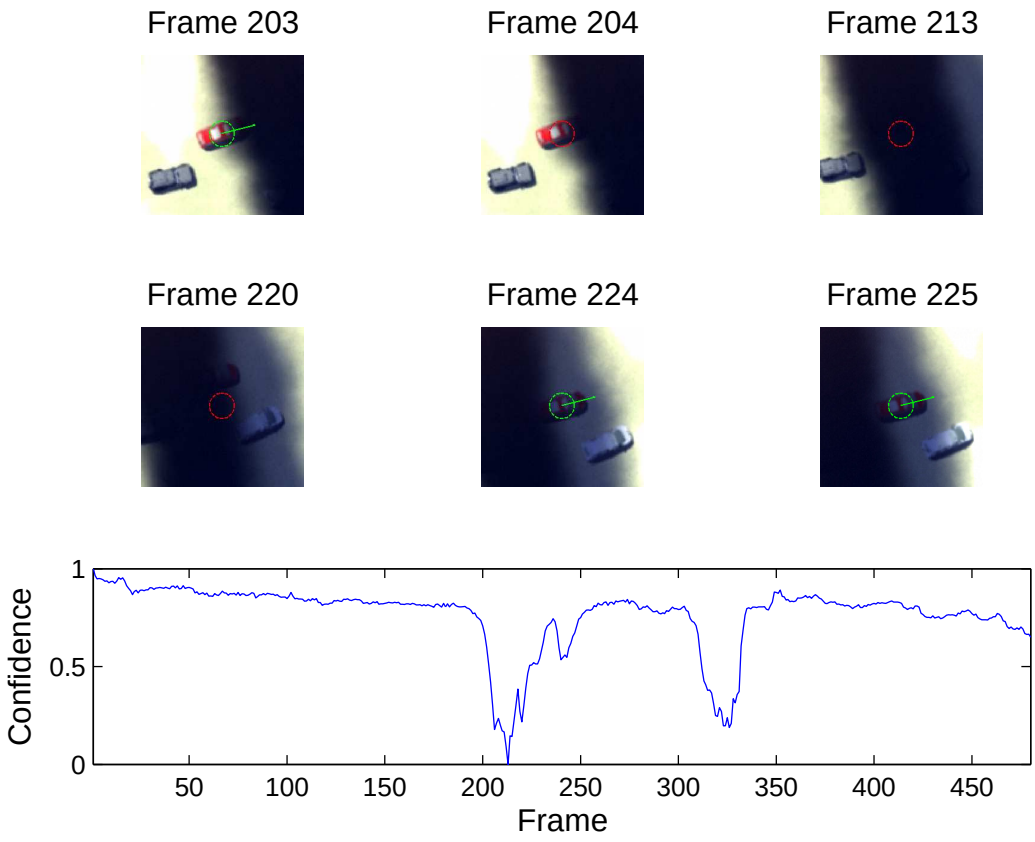

Figure 5. Frame 203: Correlation score close to decreasing below threshold value; 204: threshold breached, seacrh window extended, and extrpolation invoked; 213: full occlusion; 220: score increases and approaches threshold; 224: score becomes greater than threshold and target is reacquired; 225 window narrows and extrapolation terminated.

register a higher score, and the algorithm will begin to follow false positives. Both of these types of error are fatal.

B. Bayesian Filtering

A probabilistic framework for the video tracking problem is required to model the target's real world coordinates (x, y), velocity $(\dot{x}, \dot{y})$, and orientation θ. Furthermore, since the target of interest might be fully or partially occluded due to buildings or other visual occlusions such as smoke, a variable V is introduced to model visibility. The joint state at time t is given by $X_t = [x\ \dot{x}\ y\ \dot{y}\ \theta\ V]$. Assuming a Markov state transition, the standard state update and prediction equations are given by

$$p(X_t|Z_{1:t}) = \frac{p(z_t|X_t)p(X_t|Z_{1:t-1})}{p(z_t|Z_{1:t-1})}\ , \tag{7}$$

with

$$p(X_t|Z_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})\,\mathrm{d}X_{t-1} \tag{8}$$

and where $Z_{1:t} = [z_1\ \cdots\ z_m\ \cdots z_t]$ and z_m describe all the observations collected at time m.

C. Dynamical Models

We write the transition probability model $p(X_t|X_{t-1})$ as

$$p(X_t|X_{t-1}) = p(S_t|S_{t-1})p(\theta_t|\theta_{t-1})p(V_t|V_{t-1}), \tag{9}$$

where $S_t = [x\ \dot{x}\ y\ \dot{y}]$, θ_t and V_t are modeled to be independent of each other. Although the orientation θ_t is partially dependent on the target's position and velocity S_t and S_{t-1}, we forgo this restriction. For the target dynamic, we will use the discrete time equivalent of the near constant velocity model [12], given by

$$S_t = F \times S_{t-1} + w_t \tag{10}$$

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

where w_t is a Gaussian noise with Q_w covariance given by

$$Q_w = \begin{bmatrix} \frac{T^3}{3}S_x & \frac{T^2}{2}S_x & 0 & 0 \\ \frac{T^2}{2}S_x & TS_x & 0 & 0 \\ 0 & 0 & \frac{T^3}{3}S_y & \frac{T^2}{2}S_y \\ 0 & 0 & \frac{T^2}{2}S_y & TS_y \end{bmatrix} \tag{12}$$

That is $S_t|S_{t-1} \sim N(FS_{t-1}, Q_w)$. At time t, the orientation θ_t models the orientation of the target, with respect to the original template. This provides us with the flexibility to track changes in the image plane as the vehicle rotates in plane. This is modeled as a random walk, viz.

$$\theta_t|\theta_{t-1} \sim N(\theta_{t-1}, S_\theta) \tag{13}$$

The visibility variable V_t determines if the target is visible, or if it is temporarily obscured by smoke or walls. This affects the way the likelihood term $p(z_t|X_t)$ is estimated. The target's visibility variable will be modeled as the discrete Markov chain

$$p(V_t = 0|V_{t-1} = 1) = P_{NV}, \tag{14}$$

$$p(V_t = 1|V_{t-1} = 0) = P_V, \tag{15}$$

where P_{NV} is the probability that the target becomes occluded and P_V is the probability that the target becomes visible.

D. Observation Model

The observation model includes a non-linear mapping $H(\cdot)$ from real world coordinates to the sensor image plane. It must also accommodate events where the target is partially or fully occluded from the sensor view. We have

$$p(z_t|X_t) = \begin{cases} \exp(k \times \Pi(H(x_t, y_t), \theta_t)) & \text{if } V_t = 1; \\ \exp(NV_c) & \text{Otherwise.} \end{cases} \tag{16}$$

Here, k is scaling factor, Π is the output of the polar matching function, and NV_c is chosen such that it is, on average, higher than the background correlation score compared with the image template, but much less than the self correlation score of the image template. This gives the tracking algorithm the ability to switch to an occluded state. The exponential form of the likelihood term is used because it gives more emphasis on the larger values of the correlation function Π. Other forms were also attempted, such as flooring the negative correlation score to zero. Other, potentially better forms of the likelihood function, in the presence of noise, is discussed in [15] but this will be the subject of further work.

E. Particles Filter Algorithm

The filtering distribution of the dynamical and observation probability model above is complex and non-linear. Sequential Monte Carlo methods such as particle filters can be used to compute the inference. The key idea is to represent the required posterior density function by a set of random samples (or particles) with associated weights and to compute estimates based on these samples and weights. These particles are then propagated through time to give predictions of the posterior distribution function at future time steps. As the number of samples becomes very large, this monte-carlo characterisation becomes an equivalent representation to the usual functional description of the posterior density function. Consider

$$w_{t,p} = w_{t-1,p} \times \frac{p(z_t|X_{t,p})p(X_{t,p}|X_{t-1,p})}{q(X_{t,p}|X_{t-1,p}, z_t)}, \tag{17}$$

then the posterior filtered density is approximated by

$$p(X_t|Z_{1:t}) \approx \sum_{p=1}^{N} w_{t,p}\delta(X_t - X_{t,p}). \tag{18}$$

The choice of the importance density $q(X_{t,p}|X_{t-1,p}, z_t)$ is one of the most important issues in particle filter design. It can be shown that the optimal importance density, in the sense of minimizing the variance of the importance weights, conditioned upon $X_{t-1,p}$ and z_t, is $p(X_{t,p}|X_{t-1,p}, z_t)$ [4]. There are other suboptimal choices. One of the most popular uses the prior model density $p(X_{t,p}|X_{t-1,p})$. When substituted into equation 17, we obtain

$$w_{t,p} = w_{t-1,p} \times p(z_t|X_{t,p}). \tag{19}$$

The simple and general algorithm above forms the basis of most particle filters. It will, however, cause the variance of the importance weights to increase over time [4]. This will adversely affect the accuracy and lead to the degeneracy problem where, after a certain number of recursive steps, all but one particle will have negligible normalized weights. Consequently, this will result in a large computational effort to update particles whose contribution to the approximation of $p(s_t|Z_t)$ is almost zero. A practical measure of the degeneracy of the particle weights is the effective sample size N_{eff}, introduced in [11], namely

$$\hat{N}_{eff} = \frac{1}{\sum_{p=1}^{N} {w_{t,p}}^2}. \tag{20}$$

It is easy to see that $1 \leq N_{eff} \leq N$. A small N_{eff} indicates a degeneracy problem. When a degeneracy problem occurs (for example when N_{eff} drops below some threshold N_{thr}) a resampling step [6] has to be performed. Resampling eliminates samples with low weights and multiplies samples with high importance weights. This involves mapping a random measure $\{S_{t,p}, w_{t,p}\}_{p=1}^{N}$ into a random measure $\{S_{t,p}, \frac{1}{N}\}_{p=1}^{N}$ with uniform weights. There are several methods available for doing the remapping. The first introduction of resampling [6] was based on a simple random sampling of the particles based on the weights. However, a complete random selection is not necessary and it increases the Monte Carlo variation of the particles. Other methods such as stratified sampling [9] and residual sampling [13] may be applied. Systematic resampling [9] is another efficient method. It is simple to implement, it has a computational complexity of $O(N)$, and it minimizes the MC variation. As such, we make use of the Sampling-Importance Sampling-Resampling (SIR) filter to perform the inference and use the prior $p(X_t|X_{t-1})$ as the importance function.

IV. Simulations and Results

The maximum likelihood and particle filter methods were applied to a synthetic sequence of aerial imagery, provided by the Defence Technology Centre. The maximum likelihood confidence threshold was set to 0.7, and the particle tracking parameters are given in Table I. Figure 7 shows a time series

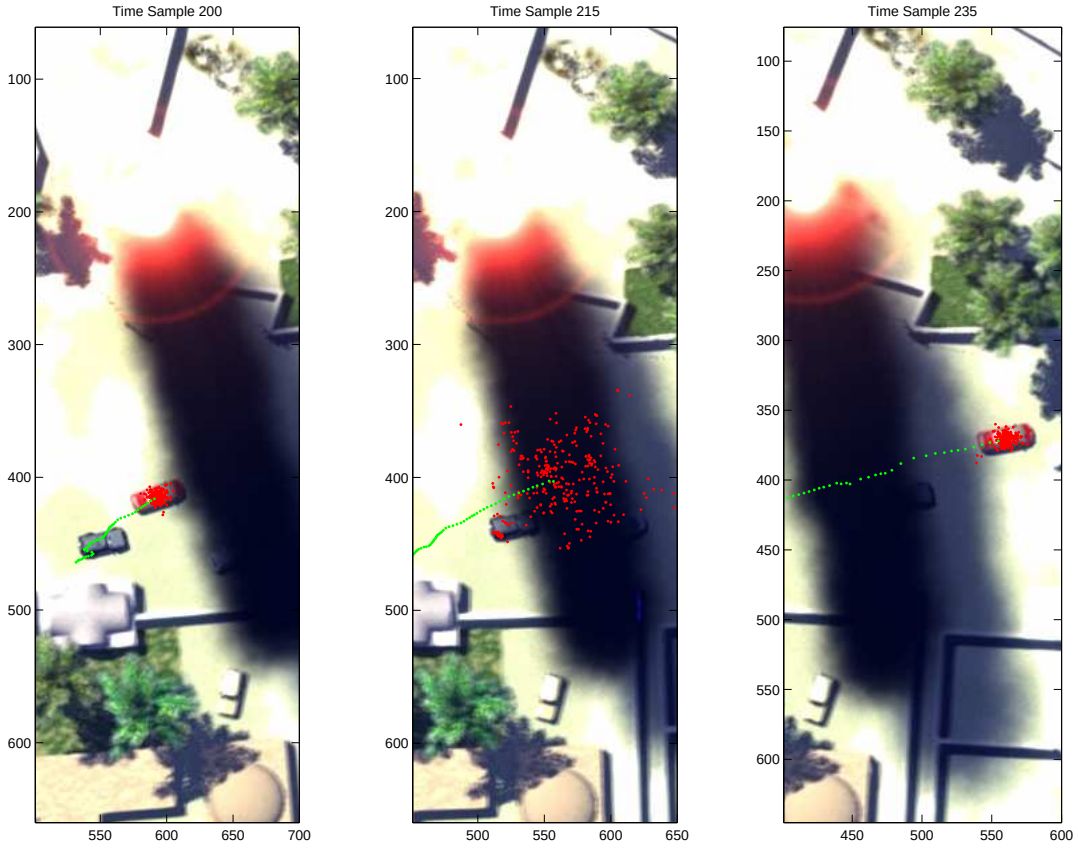

Figure 6. A sequential series of frames showing the distribution of particles (in red) and estimated track (in green) as the target enters and emerges from the thick smoke occlusion. The posterior distribution of the vehicle's position increases significantly as it enters the occlusion.

Algorithm Parameter	**Symbol**	**Value**
Time interval between measurements	T	$\frac{1}{30}$ seconds
Number of Particles		300
Motion variance	S_x, S_y	1600
Likelihood Scale	k	5
Probability of becoming occluded	P_{NV}	0.1
Probability of becoming visible	P_V	0.25
Liklihood Constant (when occluded)	NV_C	0.2

Table I
TRACKING PARAMETERS FOR PARTICLE FILTER

plot of the error $\|\mathbf{r} - \hat{\mathbf{r}}\|_2$ between the estimated $\mathbf{r}$, and true $\mathbf{r}$ position. It also shows how the confidence of the maximum likelihood, and visibility of the particle filter vary over the image sequence. The two methods result in similar performance. The small error spikes in the particle filter method occur as the posterior distribution stretches or split into two modes to account for uncertainty in the position of the target. However, the particle filter does seem to be more reliable in the event of total occlusion by smoke, around frame 210, and partial occlusion by buildings, around frame 320. The scatter plot in 8 shows error with respect to confidence/visibility for the maximum likelihood and particle filter methods. Ideally, the points should approximate some kind of monotonically decreasing function. The particle filter seems to satisfy this better than the maximum likelihood method. Figure 6 shows the distribution of the particles as the vehicle enters and emerges from the smoke occlusion. The posterior distribution of the vehicle's position increases significantly as it becomes occluded from view.

While the windowed ML search is capable of finding very good position, it is very sensitive to the threshold value. This has to be tuned carefully in order to achieve good results. The particle filter tracker is more robust to local modes of the correlation surfaces of the Polar Matching Algorithm. When the target becomes partially occluded by buildings the particle cloud bifurcates. This implies the presence of two possible modes in the posterior distribution of the position of the target.

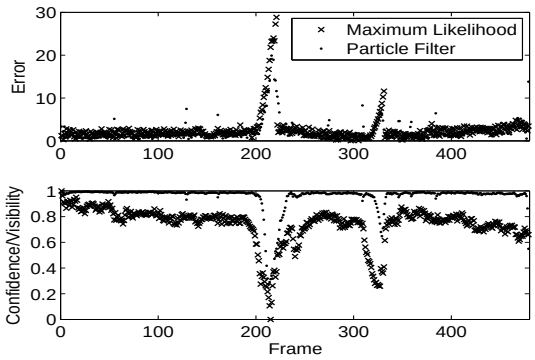

Figure 7. Error and confidence/visibility for the maximum likelihood and particle filter methods for each frame in the image sequence. The two methods result in similar performance. The spikes in the particle filter method occur as the posterior distribution stretches or split into two modes to account for uncertainty in the position of the target. However, the particle filter does seem to be more reliable in the event of total occlusion by smoke, around frame 210, and partial occlusion by buildings, around frame 320.

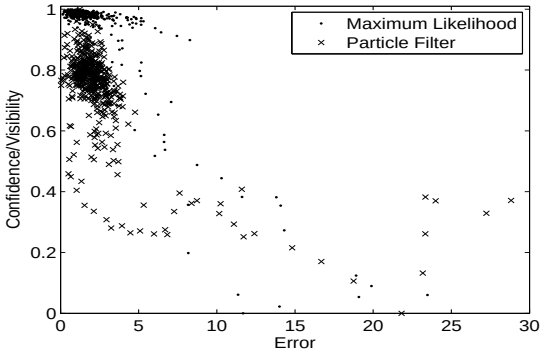

Figure 8. Scatter plot of error against confidence/visibility for the maximum likelihood and particle filter methods.

The particle cloud subsequently converges back to the true target after a few video frames. This is because the video data shows that the moving target is a more likely scenario than a static building. This also explains the small spikes in the error in Figure 7 as the posterior distribution stretches or split into two modes to account for uncertainty in the position of the target.

V. Conclusion

In this paper, we have shown that the combination of the rotation invariant dual-tree complex wavelet polar matching descriptor and the particle filter can be an effective approach to detect and track ground based targets from aerial sensor data. Polar matching offers target detection confidence scores for each position and orientation to the particle filter. The dynamic model provided by the particle filter then enhances the robustness of the polar matching and detection. Further work will include Rao-Blackwellising the particle filter to make it more efficient and the development of a more principled form of the observation model using a Track-Before-Detect approach.

VI. Acknowledgments

This research was sponsored by the Data and Information Fusion Defence Technology Centre, UK, under the Applied Multi-Dimensional Fusion and Tracking clusters. The authors thank these parties for funding this work with special thanks to General Dynamics for providing the synthetic data.

References

[1] Y. Bar-Shalom and W. D. Blair, editors, *Multitarget-Multisensor Tracking: Applications and Advances*, volume III, Artech House, 685 Canton Street, Norwood, MA 02062, 2000.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 685 Canton Street, Norwood, MA 02062, 1999.

[3] R. R. Coifman and D. Donoho. Wavelets and Statistics, Lecture notes in Statistics. In: *Translation-invariant de-noising. Antoniadis, A. and Oppenheim, G. (eds.)*, pages 125–150. Springer, 1995.

[4] A. Doucet, S. Godsill, and C. Andrieu, On Sequential Monte Carlo Sampling Methods for Bayesian Filtering, *Statistics and Computing*, 10, 2000, pp. 197–208.

[5] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, 1996.

[6] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, *Radar and Signal Processing, IEE Proceedings F*, 140, April 1993, pp. 107–113.

[7] N. G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis*, 10(3):234–253, 2001.

[8] N. G. Kingsbury, Rotation-Invariant Local Feature Matching with Complex Wavelets, *Proc. European Conference on Signal Processing (EUSIPCO)*, September 2006.

[9] G. Kitagawa, Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *Journal of Computational and Graphical Statistics*, 5, 1996, pp. 1–25.

[10] P. R. Hill, D. R. Bull, and C. N. Canagarajah. Rotationally invariant texture features using the dual-tree complex wavelet tranform. In *Proc. ICIP*, 2006.

[11] A. Kong, J. Liu, and W. Wong, Sequential Imputation and Bayesian Missing Data Problems, *J. Americian Statistical Association*, 1994, pp. 278–288.

[12] X. R. Li and V. P. Jilkov, Survey of Maneuvering Target Tracking. Part I: Dynamic Models, *IEEE Transactions on Aerospace and Electronic Systems*, 39, 2003, pp. 1333–1364.

[13] J. S. Liu and R. Chen, Sequential Monte Carlo Methods for Dynamic Systems, *J. Americian Statistical Association*, 93, 1998, pp. 1032–1044.

[14] D. G. Lowe, Distinctive Image Features from Scale-invariant Keypoints, *International Journal of Computer Vision*, 60, 2004, pp. 1333–1364.

[15] S. Maskell, A Bayesian Approach to Fusing Uncertain, Imprecise and Conflicting Information, *Infomation, Fusion*, 2007.

[16] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods - Second Edition*, Springer, New York, 2004.

[17] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

[18] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans. Inform. Theory*, 38:587–607, 1992.