VIDEO TRACKING USING DUAL-TREE WAVELET POLAR MATCHING AND PARTICLE FILTERING

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

Keywords: Dual-Tree Wavelet,Polar Matching,Video Tracking, Particle Filtering

Abstract

In this paper, we describe a video tracking application using the dual-tree polar matching algorithm. The models are specified in a probabilistic setting, and a particle filter is used to perform the sequential inference. Computer simulations demonstrate the ability of the algorithm to track a simulated video moving target in an urban environment with occasional occlusions.

1. Introduction

Detection and tracking of a known target in video sequences is a common and important problem in image processing. In this abstract we will focus on the scenario of an unmanned air vehicle (UAV) platform based image sensor as it attempts to track a ground vehicle traversing a cluttered urban environment. The objective is to estimate the position of the vehicle, in grid coordinates, to the best precision and highest time fidelity possible.

As the location of the UAV and target vary, and as the 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 have robustness or invariance to spatial shifts, rotations, and scale variations.

With this in mind, the descriptor and matching technique afforded by rotation-invariant polar matching with dual-tree complex wavelet transforms (DTCWT) recently developed by Kingsbury in 2006 [6] is adapted here, for the first time, to the task of detection. 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.

Many approaches have been proposed to tackle the problem of target tracking. These range from Kalman filter and its non-linear extensions to JPDAF trackers [1][2]. With the parallel advances in modern computational power and the developments in optimal non-linear techniques such as particle filters [5][3] and Markov Chain Monte Carlo (MCMC) [14][4], it is now possible to consider exploiting other information (such as non-linear measurement process) and potentially resulting in significant performance gains.

The detection output of the polar matching method can be fed into a tracking filter to provide smooth estimates of the target's position. However, an optimal linear filter 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 nature of the video data. To overcome these issues, we have designed a particle filter to perform the tracking.

The paper is organised as follows. Section 2 presents rotation-invariant dual-tree complex wavelet polar matching. Section 3 describe the probabilistic state-space model. Section 4 and 5 describe the dynamic models and observation model respectively. Section 6 describes the particle filter algorithm. Simulation results are shown in Section 7, followed by conclusions in Section 8.

2. Polar matching

Extending his work on the shift-invariant dual-tree complex wavelet transform [7], Kingsbury recently introduced the rotation-invariant polar matching method [6]. Owing to low redundancy, the DTCWT descriptor is more efficient than the existing popular scale- and rotation-invariant methods of SIFT [12] and Simoncelli's steerable pyramids [15]. It is adapted here to provide image matching between a small template and a larger image rather than matching keypoints of two similarly sized images, as previously reported.

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

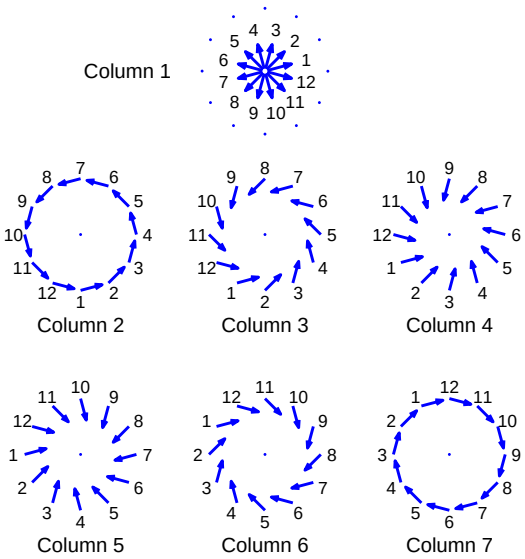

Figure 1. Locations and orientations of the DTCWT coefficients. Each orientation describes a coefficient, or conjugate, of one of the six subbands. The column numbers indicate the column of the P-matrix in which the coefficients are placed, and the numbers displayed around each circle indicate the row of the P-matrix that each coefficient is placed. Taken from [6].

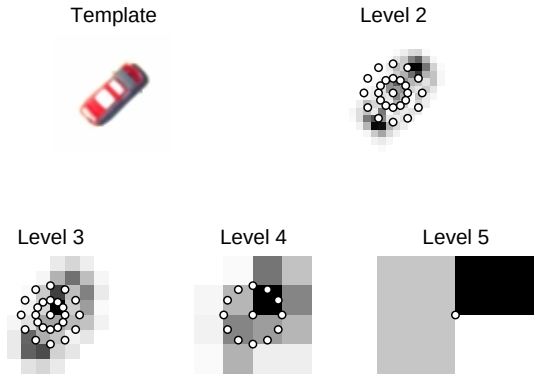

Figure 2. DTCWT coefficients of a template. The white dots indicate the locations of the coefficients that are used in the P-matrix template. Four different scales, or levels, are used. Only one of one of the 6 subbands is shown.

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.

We shall present a discussion of how to place this local feature matching method into the setting of video tracking. Two approaches will be discussed below. A simple windowed search 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. In order to cope with partial and full occlusion, the following heuristic was implemented.

for t **do**
  Compute C_t over neighbourhood $N_{t-1} \bigcup \Theta_{t-1}$
  $(x_t, y_t; \theta_t) \leftarrow \text{argmax} C_t(x,y,\theta)$
  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

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.

The problem here is that the choice of the threshold can be critical. If it is set too high, then legitimate targets may be ignored. The longer this happens, the more out of date that the extrapolation becomes. On the other hand, if the threshold is set too low, then, when the target becomes occluded, other nearby objects will register a higher score, and the algorithm will begin to follow false positives.

3. Bayesian Filtering

In this paper, we will be tracking a single moving target.

We first develop a probabilistic framework for the video tracking problem. We are interested in the target's position (x,y) and velocity $(\dot{x},\dot{y})$ in real world coordinates, as well as the orientation of the image template, θ, with respect to each video frame. Furthermore, the target of interest might be fully or partially occluded due to buildings or other visual occlusions such as smoke. Hence, we introduce a visibility variable V to model that. Hence, the joint state at time t is given by $X_t = [x\ \dot{x}\ y\ \dot{y}\ \theta\ V]$.

Assuming a Markovian state transition, the standard state update and prediction equations are given by

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

$$p(X_t|Z_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})\,\mathrm{d}X_{t-1} \tag{2}$$

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

4. Dynamical Models

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

$$\begin{aligned} & p(X_t|X_{t-1}) \\ = \ & p(S_t|S_{t-1})p(\theta_t|\theta_{t-1})p(V_t|V_{t-1}) \end{aligned} \tag{3}$$

where $S_t = [x\ \dot{x}\ y\ \dot{y}]$.

S_t, θ_t and V_t are modeled to be independent of each other. We can also make the orientation V_t to be partially dependent on the target's position and velocity S_t and S_{t-1}. Here we use a simpler model.

For the target dynamic, we will use the discrete time equivalent of the near constant velocity model [10]. This is given by

$$S_t = F \times S_{t-1} + w_t \tag{4}$$

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

w_t is a Gaussian noise with Q_w covariance given by

$$Q_w = \begin{bmatrix} \frac{T^3}{3}S_x & \frac{T^2}{2}S_x & 0 & 0 \\ \frac{T^2}{2}S_x & TS_x & 0 & 0 \\ 0 & 0 & \frac{T^3}{3}S_y & \frac{T^2}{2}S_y \\ 0 & 0 & \frac{T^2}{2}S_y & TS_y \end{bmatrix} \tag{6}$$

In other words,

$$S_t|S_{t-1} \sim N(FS_{t-1}, Q_w) \tag{7}$$

The orientation θ_t models the changes in the orientation of the video image at time t as compared with the template used. 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.

$$\theta_t|\theta_{t-1} \sim N(\theta_{t-1}, S_\theta) \tag{8}$$

The visibility variable V_t tries to determine 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 a discrete Markov chain,

$$p(V_t = 0|V_{t-1} = 1) = P_{NV} \tag{9}$$

$$p(V_t = 1|V_{t-1} = 0) = P_V \tag{10}$$

5. Observation Model

The observation of the tracking problem consists of a sequence of video images which is roughly centered on the target of interest. There is a non-linear mapping $H(\cdot)$ of the coordinates from the position of the sensor and target's real world position, to the image plane of the sensor. We will also need to consider when the target itself might be partially or fully occluded from sensor view.

$$p(z_t|X_t) = \begin{cases} \exp(k \times Polar(H(x_t,y_t),\theta_t)) & \text{if } V_t = 1; \\ \exp(NV_c) & \text{Otherwise.} \end{cases} \tag{11}$$

Here, k is scaling factor. NV_c is chosen such that it is on the 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 $Polar(\cdots)$. Other forms have also been experimented with, including flooring the negative correlation score to zero. In [13], more discussions have been provided to an optimal form of likelihood function consisting of a linear mapping in the presence of noise. This is being further investigated in the context of the Polar-Matching method.

6. Particles Filter Algorithm

The filtering distribution of the dynamical and observations probability model above is complex and non-linear. Sequential Monte Carlo methods such as particle filter can be used to do 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 characterization becomes an equivalent representation to the usual functional description of the posterior density function.

$$w_{t,p} = w_{t-1,p} \times \frac{p(z_t|X_{t,p})p(X_{t,p}|X_{t-1,p})}{q(X_{t,p}|X_{t-1,p},z_t)} \tag{12}$$

The posterior filtered density is approximated by

$$p(X_t|Z_{1:t}) \approx \sum_{p=1}^{N} w_{t,p}\delta(X_t - X_{t,p}) \tag{13}$$

The choice of the importance density $q(X_{t,p}|X_{t-1,p},z_t)$ is one of the most important issue in designing a particle filter. 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 $p(X_{t,p}|X_{t-1,p},z_t)$ [3].

There are other suboptimal choices. One of the most popular one uses the prior model density $p(X_{t,p}|X_{t-1,p})$. When substituted into equation 12, we obtain

$$w_{t,p} = w_{t-1,p} \times p(z_t|X_{t,p}) \tag{14}$$

The simple and general algorithm above forms the basis of most particle filters. However, the algorithm above will result in the variance of the importance weights increasing over time [[3]]. 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. This will result in a large computational effort devoted to updating particles whose contribution to the approximation of $p(s_t|Z_t)$ is almost zero. A practical measure of the degeneracy of the particles' weights is the effective sample size N_{eff} introduced in [9]:

$$\hat{N}_{eff} = \frac{1}{\sum_{p=1}^{N} {w_{t,p}}^2} \tag{15}$$

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 step called resampling [5] has to be performed. Resampling eliminates sample with low weights and multiplies samples with high importance weights. It basically 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 this remapping. The first introduction of resampling [5] is 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 [8] and residual sampling [11] may be applied. Systematic resampling [8] is another efficient method. It is simple to implement, its computational complexity is $O(N)$ and it minimizes the MC variation.

In this paper, we will make use of the Sampling-Importance Sampling-Resampling (SIR) filter to perform the inference. We will use the prior $p(X_t|X_{t-1})$ as the importance function.

7. Simulations and Results

We have applied the tracking filter to track a vehicle in a set of high fidelity synthetic video sequences.

We will show that the vehicle can be tracked consistently, even when it is fully occluded for a short period of time. Figure 3 shows the tracking results for the vehicle as it emerges from a full occlusion due to thick smoke. The drop in visibility can be seen in Panel (e) in Figure 3. Figure 4 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.

8. Future Works and Conclusions

In this paper, we have shown that the combination of the rotation invariant dual-tree complex wavelet polar match-

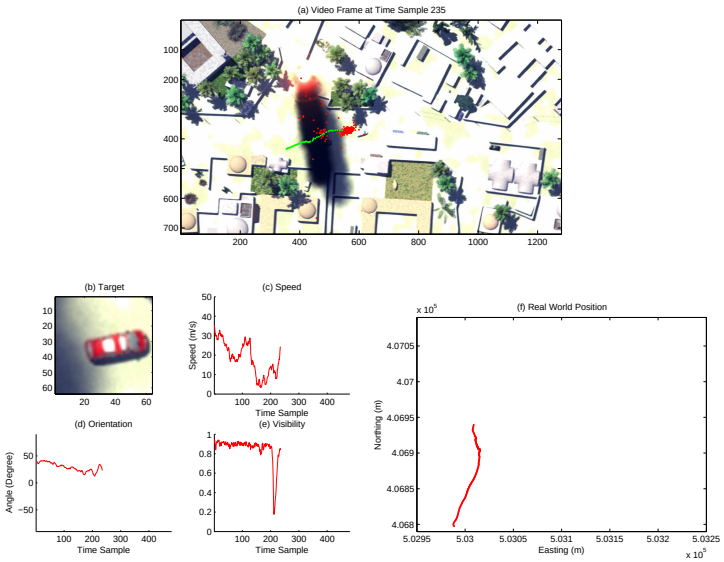

Figure 3. Tracking results at Time Sample 235. This shows the target emerging from a full occlusion due to thick smoke. Panel (a) shows the estimated track (in green) and the posterior distribution of the particles (in red) for the position of the target in the video frame. Because the target is occluded, the uncertainty of the position of the target increased significantly. Panel (b) shows the enlarged view of the target based on the current estimated position. Panel (c), (d) and (e) shows the estimated speed, orientation and visibility of the target respectively. Panel (e) shows clearly the decrease and subsequent increase in visibility as the target enter and emerge from the smoke occlusion. Panel (f) shows a plot of the real world position of the target.

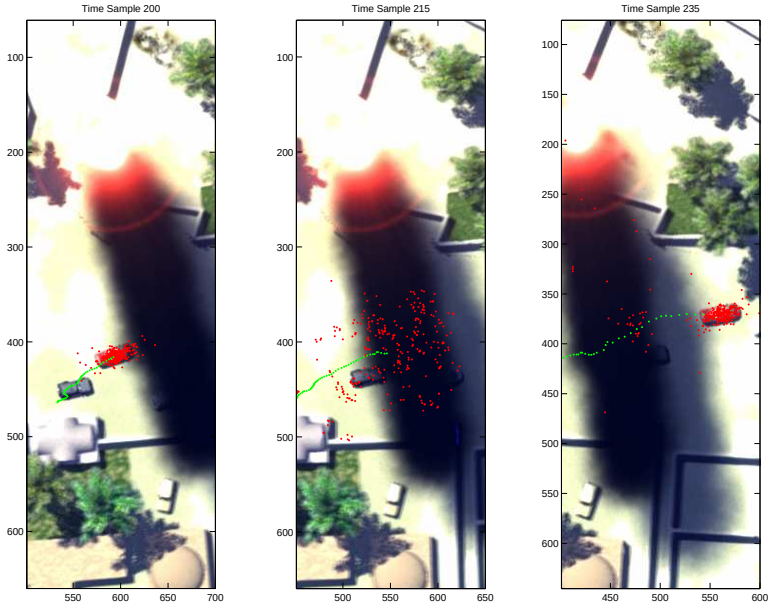

Figure 4. A sequential series of frames showing the distribution of particles (in red) and estimated track (in green) as the target enters and emerges from the thick smoke occlusion. The posterior distribution of the vehicle's position increases significantly as it enters the occlusion.

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 1. Tracking Parameters for Particle Filter

ing descriptor and the particle filter can be an effective approach to detect and track ground based targets from UAV 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.

9. Acknowledgments

The research of XXXXX was sponsored by the Data and Information Fusion Defence Technology Centre, UK, under the Tracking Cluster. The authors thank these parties for funding this work. The authors will also like thank General Dynamics for providing the high fidelity synthetic data. The authors are grateful to Simon Maskell for the discussions on the form of likelihood function.

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

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

[5] 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.

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

[7] N. G. Kinsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals", *Journal of Applied and Computational Harmonic Analysis*, 10, May 2001, pp. 234–253.

[8] 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.

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

[10] 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. 13331364.

[11] J. S. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems", *J. Americian Statistical Association*, 93, 1998, pp. 10321044.

[12] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints", *International Journal of Computer Vision*, 60, 2004, pp. 13331364.

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

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

[15] E. P. Simoncelli and W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation", *Proc. ICIP*, 3, October 1995, pp. 444–447.