Tracking of motor vehicles from aerial video imagery using the OT-MACH correlation filter

Bangalore Nagachetan, Young Rupert, Birch Philip, Chatwin Christopher

Department of Engineering and Design, School of Engineering and Informatics, University of Sussex, Brighton BN1 9QT
Phone: 44 (0)1273 678908, e-mail: r.c.d.young@sussex.ac.uk

ABSTRACT

Accurately tracking moving targets in a complex scene involving moving cameras, occlusions and targets embedded in noise is a very active research area in computer vision. In this paper, an optimal trade-off maximum correlation height (OT-MACH) filter has been designed and implemented as a robust tracker. The algorithm allows selection of different objects as a target, based on the operator's requirements. The user interface is designed so as to allow the selection of a different target for tracking at any time. The filter is updated, at a frequency selected by the user, which makes the filter more resistant to progressive changes in the object's orientation and scale. The tracker has been tested on both colour visible band as well as infra-red band video sequences acquired from the air by the Sussex County police helicopter. Initial testing has demonstrated the ability of the filter to maintain a stable track on vehicles despite changes of scale, orientation and lighting and the ability to re-acquire the track after short losses due to the vehicle passing behind occlusions.

Keywords: OT-MACH filter; correlation; active contours; vehicle tracking; aerial video

1. INTRODUCTION

Accurately tracking moving targets in a complex scene involving moving cameras, occlusions and targets embedded in noise is a very active research area in computer vision. An optimal trade-off maximum correlation height (OT-MACH) filter has been designed and implemented as a robust tracker [1-6]. The tracker has been tested both on colour visible band as well as infra-red band video sequences acquired from the air by the Sussex police helicopter. An interrupt based user interface which is active for the whole duration of tracking in the videos or live feed has been developed. The user defines a circular area around the target using an interrupt to start tracking the selected target. The area of support for extracting training images of the target has been developed using: rectangular, circular and active contour based methods for training image extraction. Cross-correlation of the filter function with subsequent video frames is parameterised via a user initialised configuration file. The algorithm allows selection of different objects as a target, based on the operator's requirements. The user interface is designed so as to allow the user to select a different target for tracking at any time. The filter is updated at a frequency selected by the user.

The filter parameters are initialised and amended using an initialisation text file. The initialisation file is used to fetch the frequency of up-date, i.e. rate of correlation, and filter parameter values provided by the user. The design makes the filter more resistant to progressive changes in the object's orientation and scale. The tracker has been tested on both colour visible band as well as infra-red band video sequences acquired from the air by the Sussex County police helicopter. Initial testing has demonstrated the ability of the filter to maintain a stable track on vehicles despite changes of scale, orientation and lighting and the ability to re-acquire the track after short losses due to the vehicle passing behind occlusions.

2. INTERRUPT USER INTERFACE

A user interface has been developed to select the target from a video source. The user interface is activated once the user finds a visual target in the scene by pressing any key from the keyboard. The interface creates a mouse interaction protocol on the current frame. The interface takes a snap shot of the current frame and displays the image for user target selection. With the assistance of the mouse, the user can draw and drag a circle over the target to be selected for tracking. The initialised parameters such as the x and y co-ordinates of the centre and the diameter of the circle are stored for further use to allow real-time fabrication of the filter. The circular area around the target is traced to find the coordinate vector at the circumference of the circle. These coordinate points are used to extract the reference image from the target. The extracted circular target is then passed on to one of the supporting methods namely: rectangular, circular and active contour based extraction, to process the reference image selected from the target. The processed target is automatically cropped and used for training the filter. Figure 1 below illustrates an original frame in which the initialised circle has been drawn over the target using the developed interface.

Figure 1: Initialisation circle over the target vehicle to be tracked

3. REFERENCE IMAGE EXTRACTION

3.1 Rectangular target extraction

The rectangle coordinate is obtained from the initialisation step by computing a bounding rectangle around the selected circle. The rectangle parameters (x_1, y_1) and (x_2, y_2) are computed and stored in the configuration file. Width and height are calculated as the difference of the parameters stored. A blank reference template image of frame size is created. The centre of the frame is computed as frame_width/2 and frame_height/2 for x and y coordinates respectively. The current frame is cropped at the initialised rectangle and copied into the blank image at the centre. The starting point for copying the target to the centre is straightforwardly calculated as:

$$(X_c, Y_c) = (\text{Frame\_width}/2, \text{Frame\_height}/2) \tag{1}$$

where the centre of the blank frame is denoted as (X_c, Y_c). The co-ordinates used to copy the cropped target into a blank frame are, clearly, then given as:

$$(X, Y)_{(\text{to copy})} = (X_c - (\text{rectangle\_width}/2), Y_c - (\text{rectangle\_height}/2)) \tag{2}$$

The windowed target, shown in Figure 2, can then be used to train the filter. The windowed reference is rotated in 2 degree increments to +6 and -6 degrees, thus obtaining seven reference images to multiplex into the filter function. This ensures some degree of in-plane rotation tolerance of the filter and facilitates its ability to maintain a track on the vehicle for n frames, after which the filter is up-dated.

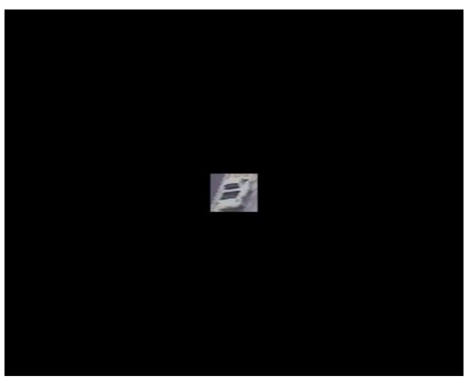

Figure 2: Target reference image used to train the filter

3.2 Circular target extraction

The circle coordinates are computed from the circle selected on the target. The centre of the circle (x, y) is obtained from the initialisation step. The circumference (C) is obtained by tracing the circle. The centre of the frame is computed as frame_width/2 and frame_height/2 for x and y coordinates, respectively. The current frame is cropped at the initialised circle and copied into the blank image at the centre. The starting point for copying the target to the centre is calculated according to Equation (1).

3.3 Active contour based target extraction

The Active Contour work employs the principle of energy minimization [7]. The energy of each coordinate point is calculated based on the neighbourhood pixels of each point. A difference of Gaussian (DoG) filtered image of the circular area is computed to emphasise the exterior edges of the target. The energy minimisation process is executed until the exterior edge of the target is contoured. The process is maintained and controlled by the number of iterations throughout the contour vector points. The number of iterations required and the snake parameters are also included in the configuration file for standardised usage of the algorithm by a user. The energy functional computed and iterated for each coordinate point is described by the expression in Equation (3):

$$E^{*}_{snake}(s) = E_{\text{int}}(v_s) + E_{image}(v_s) \qquad (3)$$

where * means that this is a continuously updating snake energy
This can be expressed as:

$$E^{*}_{snake} = \alpha(s).\left|\frac{dv_s}{ds}\right|^2 + \beta(s).\left|\frac{d^2 v_s}{ds^2}\right|^2 + \gamma(s)E_{edge} \qquad (4)$$

where the first-order and second order differentials are approximated for each point that is searched in the local neighbourhood of the currently selected coordinate point. The weighting parameters α, β and γ are all functions of the contour. E^{*}_{snake} is thus the overall Snake energy term and E_{edge} the computed edge energy. By calibrating the Snake, the exterior edge of the target is contoured. The edge contour gives the coordinate vector of the target in the frame. The shape vector thus obtained is used to segment and extract the target object from the scene. It is placed at the centre of a blank zero background image, as shown in Figure 3 to create training images for filter initialisation and computation [10].

The contour's coordinate vector of the exterior edge of the target is used to extract the target from the frame interrupted. A blank reference template image of frame size is created. The centre of the frame is computed as frame_width/2 and frame_height/2 for x and y coordinates, respectively. The current frame in which the target is contoured is used to extract the contoured object and is copied to the centre of the blank frame.

Figure 3: The reference image generated for training the filter

The selected target, shown in Figure 3, can then be used to train the filter. The reference image is rotated by 2 degree increments to +6 and -6 degrees, thus obtaining seven reference images to multiplex into the filter function. This ensures some degree of in-plane rotation tolerance of the filter and facilitates its ability to maintain a track on the vehicle for n frames, n being the frequency of the filter upgrade parameter fetched from the configuration file.

3.4 Computing rotationally multiplexed reference image

A double precision addition of several rotated reference templates is performed to compute a rotationally multiplexed reference image. The reference image obtained from the processing, as explained in the previous section, is rotated in increments and decrements of θ degrees. The θ value and the level of rotational multiplexing are fetched from the configuration file. For example if the θ value is 2 degrees and the level is 3, then the reference image is rotated in increments and decrements of 2 degrees between -6 and +6 degrees to obtain seven rotated reference images. The reference images are of double precision and are added to obtain a multiplexed reference image, which is further used in fabricating the OT-MACH filter. A typical example of a rotationally multiplexed reference image is shown in Figure 4 [7-9].

Figure 4: Rotationally multiplexed reference image

4. REAL-TIME IMPLEMENTATION OF THE OT-MACH FILTER

The OT-MACH filter is designed by passing the set of reference images to the filter design function. The filter is applied to every m^{th} input video frame to generate a correlation peak, the location of which indicates the position of the target vehicle in the video frame, a typical example being shown in Figure 5 below. The rate of cross-correlation peak generation is controlled by parameter m initialised in the configuration text file by the user. The target location is then displayed using cross-hair markers as shown in the figures below. The filter can be updated in real-time or changed, based on the user's requirement, by initialising the update frequency in the configuration file. The rotational multiplexing increases the tolerance of the filter to changes of vehicle rotation angle between filter up-dates. There are also progressive changes in scale, due to variations in distance of the camera from the target vehicle, but these are sufficiently small that they can be accommodated by the filter up-date process described below.

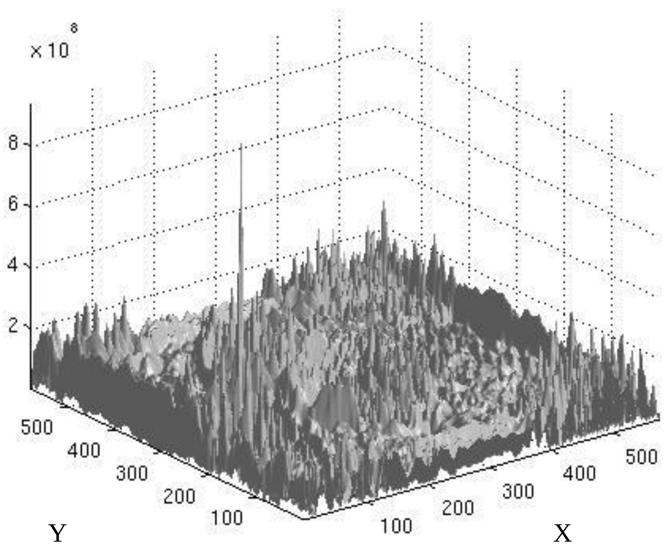

Figure 5: Correlation plane with peak location at (X, Y) = (195,342)

At every m^{th} frame (for example if m =25, then every second of the video sequence) the filter is updated. To do this, a correlation of the filter and target frame takes place and a measurement of the height of the correlation peak at the target location is made. To ensure the filter function has correctly identified the target vehicle, the correlation peak height obtained is compared to the average of the previous values at the rate of correlation set by the user in the configuration file. This allows the algorithm to take account of those cases in which the target may not be visible during the m^{th} frame cycle of measurement. Once a correlation peak of sufficient height has been obtained, the filter is up-dated using the current image. The reference image for the current frame is obtained, and a new rotationally multiplexed filter is created, as described previously. If the user decides to change the target at anytime, a keyboard interrupt can be used to switch 'on' the user interface to select a new target. This captures the current interrupted frame and allows the user to retrain the filter with a new target. The updating of the filter takes place as explained in the previous sections. The correlation of every frame at the rate interval is again carried out to maintain the track for the next m frame sequence.

The Fastest Fourier transform in the West (FFTW) is a C subroutine library for computing the discrete Fourier Transforms (DFT) in one or more dimensions [11]. Using C library linkers the FFTW has been interfaced to the C program developed for the filter design. The output from the FFTW is not a shifted FFT. Shifting the zero component of the fast Fourier transform (FFT) to the centre of the spectrum is performed using the function FFTWSHIFT() implemented in C. The FFTWSHIFT() function rearranges the output obtained from the FFTW by swapping the first quadrant with the third quadrant and the second quadrant with the fourth quadrant.

The Fourier transformed output from the FFTW routine is converted to an image in the IPL_IMAGE format of the OpenCV library. Several OpenCV library functions were used to perform the swap operations to obtain the Shifted FFT (FFTSHIFT).

5. COMPARISON OF THE OT-MACH TRACKER WITH KALMAN AND PARTICLE FILTERS

An extended Kalman filter and colour histogram based Particle filter tracker has been tested to compare with the accuracy of OT-MACH filter tracker. The results of the comparison have clearly proven the OT-MACH filter tracker as being robust and efficient in non-linear, noisy and dynamic environments. The Particle filter tracker tested is colour based and hence has not been used to verify the accuracy of OT-MACH filter tracker on infra-red video sequences.

In the video sequences (snapshot figures shown below) the Kalman filter result, signifying the predicted position of the target, is indicated with a red cross-hair marker while the OT-MACH tracker result is marked with a yellow cross-hair. The Particle filter result uses blue markers to signify the particles composing the filter along with a red track marker to trace the trail of the target. The noisy, dynamic and non-linear target scenes that we have used as test sequences have resulted in the inability of the Kalman and Particle filters to perform as accurate trackers. The OT-MACH filter tracker has some degree of in-plane rotation and scale invariance as well as tolerance to orientation changes of the target object. Due to frequent filter updates it has higher track accuracy with varying velocity and other non-linear target situations. Figure 6 and 7 show examples of tracking failure using the Kalman and Particle filter in situations where the OT-MACH filter is able to maintain successful tracking.

Figure 6: Kalman filter (Red) and OT-MACH filter tracker (Yellow)

Figure 7: Particle filter (blue, green and red) and OT-MACH filter tracker (Yellow)

6. USE OF AN EXTENDED KALMAN FILTER

There are situations when multiple target-like objects are found in a scene. If the objects are identical, then the OT-MACH filter will detect multiple objects as targets. In order to help avoid loss of the selected target and to differentiate between targets and non-targets (in situations, for instance, in which very similar road vehicles pass close to each other on a road), an extended Kalman filter has been developed; this allows a predictive track to be applied to the target object. It is implemented by estimating the uncertainty of the predicted position. A weighted average of the predicted value is then computed and most weight given to the value with least uncertainty. The maximum correlation peak location of the target is used in the filter as the original location. After obtaining about 10 coordinates of the target from the correlation result, the

Kalman filter is trained to predict approximate future locations. The Kalman filter estimated coordinate is then used as a reference point to retrain the OT-MACH filter.
However, the extended Kalman filter, although a non-linear filter, it is not an optimal estimator when dealing with noisy frames, varying velocity targets and extreme scale variations. Encouragingly, the OT-MACH tracker has been found to work accurately with varying scale, orientation and velocity of the target object. (In contrast, the original Kalman filter, as a target predictor, is feasible only for constant speed target scenes.) Several methods have been investigated to obtain a multi-target solution for robust tracking. The colour based Particle filter has been found to fail to precisely locate the target in noisy sequences and in infra-red band videos. The methods being considered as future work for improving the solution to close proximity multi-target object situations are briefly discussed in the Conclusion section.

7. RESULTS

Tests were conducted and several examples of tracking targets at different scales and orientations, a few examples being shown in Figures 8-15 below.

Figure 8: Sussex Police video (1) Frame 5 with false objects

Figure 9: Sussex Police video (1) Frame 15 with multiple false objects

Figure 10: Sussex Police video (1) Frame 240, scale changed

Figure 11: Sussex Police infra-red video frame 104, scale changed with Gausian noise

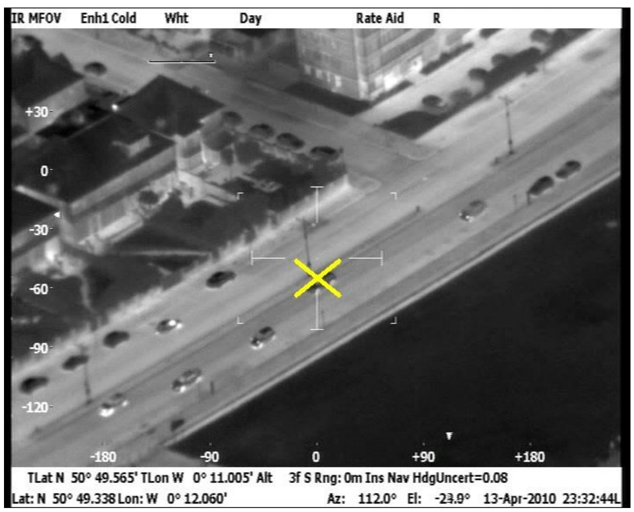

Figure 12: Sussex Police infra-red video frame 265, scale changed

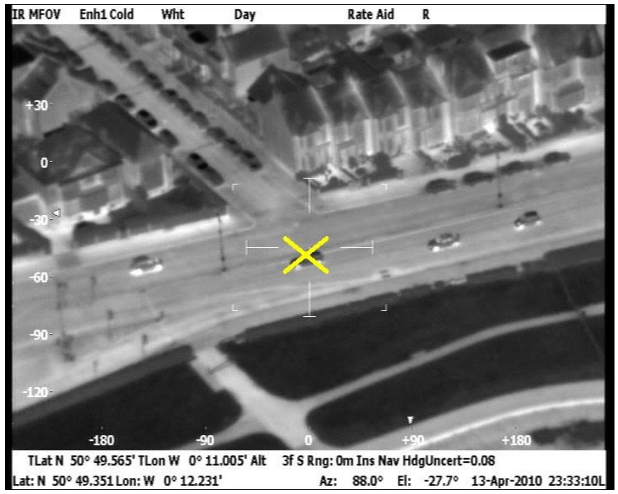

Figure 13: Sussex Police infra-red video frame 435, scale and orientation changed

Figure 14: Sussex Police video (3) frame 35

Figure 15: Sussex Police video (3) frame 548, scale and orientation changed

8. CONCLUSIONS

The OT-MACH tracker has been optimised and implemented as a robust vehicle tracker. The filter is rotation multiplexed and applied at a frame rate initialised by the user on the video sequences, with a filter up-date being implemented every mth frame. On the video sequences for which it has been tested, a few typical examples have been displayed in this paper. The algorithm can switch from using rectangular, circular or active contour based extraction methods. Compared to rectangular and circular extraction methods the active contour snake is found to allow the maintenance of a strong and accurate correlation peak at the target location. The user interface is designed so as to allow the user to interrupt and select a new target from a current frame. The OT-MACH filter is frequently updated by retraining with rotationally multiplexed reference images extracted and processed during an interval period chosen by the user. A Kalman filter predictor is implemented to conditionally solve the problem of multiple targets in the scene. The results obtained have been discussed and illustrated in section 7. From the tests performed to date, the OT-MACH tracker shows considerable promise and has the capability to perform accurately in cluttered and noisy sequences. It is found to be accurate in recognising and tracking the target, out-performing an extended Kalman filter and colour based Particle filter approach in noisy and dynamic sequences. Nevertheless, the Kalman filter is useful for disambiguation when multiple targets are in close proximity and in application to constant velocity targets; hence running the two algorithms in parallel is a useful approach to solve the problem of tracking multiple target-like objects.

Further work will consider modification of the area of support for more robustly resolving multiple target confusion. Current techniques such as the SIFT (scale invariant feature transform) and SURF (speeded up robust feature detector) algorithms for scale invariant feature recognition are also being considered for enhancing the performance of the OT-MACH tracker. Combinations of these methods may allow the target to be continuously tracked even in the presence of close proximity multiple target-like objects in the scene and will help provide a predictive track to assist the re-acquisition of the target vehicle if the track is interrupted. A new approach to train a Particle filter using the correlation plane result obtained from the OT-MACH tracker is being investigated to improve the accuracy and the ability to track in multi-target scenarios. Further study is also being conducted in the area of low pixel count object detection and tracking in low signal-to-noise ratio conditions to assist in real-time target tracking situations from an extended range.

REFERENCES

1. Mahalanobis, B.V.K. Vijaya Kumar, S. Song, S.R.F. Sims, J.F. Epperson, “Unconstrained correlation filters”, *Applied Opt*ics, Vol. 33, pp. 3751-3759, (1994).
2. A. Mahalanobis, B. V. K. Vijaya Kumar, D. Casasent, “Minimum average correlation energy filters”, *Applied Optics*, Vol. 26, No. 17, pp. 3633-3640, September, (1987).
3. B. V. K. Kumar, Minimum Variance Synthetic Discriminant Function, *Journal of Optical Society of America A3*, Vol. 3, pp. 1579-1584, (1986).
4. Z. Bahri and B. V. K. Kumar, “Generalized synthetic discriminant functions”, *Journal of Optical Society of America*, Vol. 5, No. 4, 562-571 (1988)
5. Hanying Zhou and Tien-Hsin Chao, MACH filter synthesising for detecting targets in cluttered environment for gray-scale optical correlator, *Proc. SPIE*, Vol. 715, pp. 394-398, April 1999
6. R.K. Wang, R.C.D. Young, C.R. Chatwin, Assessment of a Wiener Filter Synthetic Discriminant Function for Optical Correlation - Journal of Optics and Lasers in Engineering, Elsevier Applied Science, Vol. 22, No. 1, pg. 33-51, (1995).
7. Mark Nixon, Alberto Aguado, Feature extraction and image processing, Newenes, Elsevier, 2002.
8. P. Bone, I. Kypraios, R. C. D. Young, C. R. Chatwin, Fully invariant object recognition in cluttered scenes, Invited Paper, Proc. SPIE, Information Technologies, Editors: A. Andriesh, V. Perju, Chisinau, Moldova, May 2004.
9. P. Bone, R. C. D. Young, C. R. Chatwin, Position, rotation, scale, and orientation invariant object tracking from cluttered scenes, SPIE Defense and Security Symposium, Optical Pattern Recognition XVII, Editors: D.P. Casasent, T-H. Chao, Kissimmee, Florida, USA, SPIE, April 2006
10. P. Bone, R. C. D. Young, C. R. Chatwin, Position, rotation, scale and orientation invariant multiple object recognition from cluttered scenes, Optical Engineering, Volume 45, pp. 077203-1 to 8, No. 7, 2006
11. Nagachetan Bangalore, Rupert Young, Philip Birch, Chris Chatwin, Tracking Moving Objects Using Bandpass Filter Enhanced Localisation and Automated Initialisation of Active Contour Snakes, ICGST International Journal on Graphics, Vision and Image Processing, GVIP, Volume 10, Issue IV, pp.1 to 8, October 2010.
12. Fastest Fourier Transform in the West, Website: http://www.fftw.org/