ENHANCED ROBUST PHASE CORRELATION BASED SUB-PIXEL FEATURE MATCHINIG FOR TARGET MOTION ESTIMATION

Hongshi Yan and Jian Guo Liu

Department of Earth Science and Engineering, Imperial College London
Prince Consort Road, London SW7 2AZ, United Kingdom - (h.yan, j.g.liu)@imperial.ac.uk

KEY WORDS: Enhanced Robust Phase Correlation (ERPC), Sub-pixel Feature Matching, Moving Target Detection, Target Motion Estimation

ABSTRACT:

This paper presents an enhanced robust phase correlation (ERPC) algorithm for sub-pixel feature matching and its application in target motion estimation for aerial video surveillance. The ERPC can cope with very large motion measurement on the one hand and improve the sub-pixel accuracy by entirely avoiding the ill-posed problem of 2D phase unwrapping in 2D fitting technique of phase correlation on the other. The key advantage of EPRC is its robustness and sub-pixel accuracy which are essential for precise target speed measurement. Furthermore, EPRC is solely applied to the certain parts (not to the whole scene) of a scene where the moving targets are detected, which greatly improves robustness and computing speed of the EPRC based target motion estimation. Finally, we introduce our robust camera compensation and moving target detection scheme. With this simple scheme, we are able to efficiently estimate the motions of multiple targets at sub-pixel accuracy.

1. INTRODUCTION

An accurate estimation of target motion in an image sequence is a crucial step for target tracking in aerial video surveillance. A number of methods, such as optical flow based, correspondence based and background model based, have been proposed [Kumar *et al.*, 2001]. In this paper, we present efficient phase correlation based target motion estimation technique. Phase correlation feature matching method has been a popular choice in estimating the global or local translational motions between two similar images due to its remarkable accuracy and its robustness to uniform variations of illumination and signal noise in images. Several phase correlation methods [Stone *et al.*, 2001; Foroosh *et al.*, 2002; Hoge, 2003] have been proposed for estimating the translational shift with sub-pixel accuracy between image pairs.

However, if there are moving objects in a scene, most existing phase correlation based technique will fail to estimate the frame shift. In this paper, we propose an enhanced robust phase correlation (ERPC) technique for sub-pixel feature matching, and apply ERPC in target motion estimation for aerial video surveillance. We present that the proposed ERPC technique is capable of measuring the motion of a moving target in a stationary background as long as the target is the dominant feature in the image frame. The key advantage of the proposed technique is its robustness and its sub-pixel accuracy in target motion estimation. In addition, the ERPC algorithm is only applied to the certain parts of a scene where the moving targets are detected, which greatly improves robustness and computing speed of the phase correlation based target motion estimation.

2. ERPC BASED MOVING TARGET SPEED MEASUREMENHT

2.1 Basics of ERPC

Phase correlation provides straight-forward estimation of rigid translational motion between two images, which is based on the well-known Fourier shift property: a shift in the spatial domain of two images results in a linear phase difference in the frequency domain of the Fourier Transforms (FT). Given two 2D functions $g(x,y)$ and $h(x,y)$ representing two images related by a simple translational shift a in horizontal and b in vertical directions, and the corresponding Fourier Transforms are denoted $G(u,v)$ and $H(u,v)$. Thus,

$$H(u,v) = G(u,v)\exp\{-i(au+bv)\} \quad (1)$$

The phase correlation is defined as the normalised cross power spectrum between G and H, which is a matrix:

$$Q(u,v) = \frac{G(u,v)H(u,v)^*}{\left|G(u,v)H(u,v)^*\right|} = \exp\{-i(au+bv)\} \quad (2)$$

If $G(u,v)$ and $H(u,v)$ are continuous functions, then the inversed Fourier Transform (IFT) of $Q(u,v)$ is a delta function. The function peak identifies the integer magnitude of the shift between the pair of images [Kuglin, and Hines, 1975]. To achieve the translation estimation at sub-pixel accuracy based on the delta function of the IFT of phase correlation matrix $Q(u,v)$, a common approach is to oversample images $g(x,y)$ and $h(x,y)$ to sub-pixel level before the FT of phase correlation operations. This however will increase the computing load dramatically. Recognised the drawback, many researchers looked for a direct solution in frequency domain based on the phase correlation matrix defined in (2). As the magnitude of $Q(u,v)$ is normalised to 1, the only variable in (2) is the phase difference defined by $au+bv$, where a and b are the horizontal and vertical magnitudes of the image shift between $g(x,y)$ and $h(x,y)$. If we can solve a and b accurately based on the phase correlation matrix $Q(u,v)$, then the non-integer translation estimation at sub-pixel accuracy can be achieved without applying IFT. Such direct frequency domain approaches [Stone *et al.*, 2001; Hoge, 2003] has been proved more accurate than that based on the delta function method.

The phase difference angle $c= au+bv$ in (2) is simply a planar surface through the origin in u-v coordinates defined by coefficients a and b. Thus a complicated problem of complex numbers in frequency domain becomes a simple issue of finding the best 2D fitting of the phase difference angle data in

Q(*u*,*v*) to a plane of phase difference in the coordinates of *u* and *v*. The phase shift angle *c* is 2π wrapped in the direction defined by *a* and *b*. Any 2D fitting technique for *c* is not possible without a 2D unwrapping. However, 2D unwrapping on the phase angle data in the *Q*(*u,v*) is often unreliable due to the noisier data of *Q*(*u,v*) and results in failure of finding *a* and *b* correctly [Foroosh *et al.*, 2002; Hoge, 2003]. In our previous work, we applied a phase fringe filtering technique to reduce the noise in the periodic data of phase correlation matrix before the 2D phase unwrapping [Liu and Yan 2006]. However, this technique may become malfunctioning for dense fringes induced from large image shifts due to the restriction of the smallest fringe filter size (3×3).

Here, we propose a novel approach namely the Enhanced Robust Phase Correlation (ERPC) algorithm for image feature matching with sub-pixel accuracy, which entirely avoids the ill-posed problem of 2D phase unwrapping. The ERPC comprises two stages processing. Firstly, the image correspondence is estimated at integer pixel level accuracy using a Delta function based phase correlation matching method. Thus, the disparity estimation error becomes no greater than 1 pixel for every corresponding correlation point. Then, the disparity measurement is refined to sub-pixel accuracy for corresponding points through robust 2D fitting.

In the sub-pixel shift estimation stage, the Quick Maximum Density Power Estimator (QMDPE) [Wang and Suter 2004 A] is applied to find the best 2D fitting estimates of the phase angle data with sub-pixel shift only. As a pair of corresponding correlation points have only sub-pixel shift each other after the integer shift has already been identified and compensated in the first processing stage, the phase difference between them is within 2π and thus the phase unwrapping is no longer necessary before the robust fitting estimation. The QMPDE is regarded as the most robust fitting method. The benefit of using the QMDPE robust estimator is that the optimal fitting estimates can be obtained from the noisy phase angle dataset. Initial tests indicate that in a window based phase correlation scanning processing, the ERPC algorithm can achieve better than 1/50th pixel feature matching accuracy with small image size.

2.2 Bench Mark Test

An example in Figure 1 shows the effectiveness and the accuracy of motion estimation of a moving target in a stationary background using ERPC technique in different image size. We inserted a picture of a tank model into different positions in a sand desert image (512×512). The tank position change between Figure 1(a) and Figure 1(b) is 45 pixels (to left) horizontally and 12 pixels (down) vertically. We resample the pair of images to generate a set of image pairs with size from 256×256 to 32×32. ERPC method was applied to the set of image pairs with different image size for the translational shift estimation of the moving tank in a stationery background. The experimental results are shown in Table1, which indicate that the ERPC technique is reliable for sub-pixel translational shift estimation of the moving target in different image size, even in a image size as small as 32× 32. It should mention that the accuracy of ERPC motion estimation will drop if the moving target is too small compared with the background.

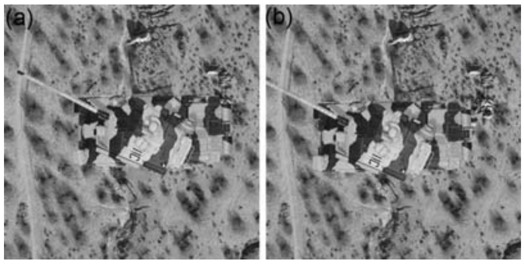

Fig. 1: An image pair (a) and (b) show a moving tank in a stationary background.

Image Size	True (*x*, *y*)	ERPC (*x*, *y*)
512×512	-45 12	-45.002 11.9957
256×256	-22.5 6	-22.5238 5.9521
128×128	-11.25 3	-11.2845 2.9626
64×64	-5.625 1.5	-5.6709 1.5411
32×32	-2.8125 0.75	-2.8053 0.7315

Table 1: Motion estimates of tank through ERPC method in different image size.

3. MOVING TARGET DETECTION

3.1 Robust Camera Motion Estimation

In aerial video surveillance, moving target detection and speed estimation are the fundamental research topics. Due to the camera motion, the background of a scene appears moving besides the target motion. The motion of actual targets must be distinguished from the global motion of the scene.

In our approach, the camera motion is compensated based on the robust feature based sparse optical flow estimation. The Kanade-Lucas-Tomasi (KLT) feature tracker [Shi and Tomasi 1994] is used to match corner features between adjacent pairs of video frames to obtain a sparse estimate of the optical flow field. The tracked features are simply used as control points to which a frame-to-frame parametric registration model is fitted by the highly robust QMDPE method [Wang and Suter 2004 A]. Most of existing methods use another robust estimator RANSAC for robustly fitting camera motion model. However, RANSAC requires the user to have priori-knowledge of the scale of inliers, which is not easy to set correctly in most cases. Furthermore RANSAC is less robust than the QMDPE as indicated in [Wang and Suter 2004 B]. The highly robust estimator QMDPE can tolerate more than 80% of outliers and it is thus used to estimate background motion from the selected sparse motion vectors.

This robust estimator employs the mean shift procedure [Comaniciu and Meer 2002] to find the local maximum density power [Wang and Suter 2004 A] $\psi_{\max}$, i.e., $\max_J \ \psi_J$, where *J* is the index of sub-samples. We use affine model to fit the sparse optical flow field. Let (u_i, v_i) be the velocity vector at feature point (x_i, y_i). The fitted affine model residuals of the point (x_i, y_i) can be written as

$$r_u = u_i - (a_0 + a_1 x_i + a_2 y_i)$$
$$r_v = v_i - (a_3 + a_4 x_i + a_5 y_i) \quad (1)$$

Let ψ_u and ψ_v be the density power functions of r_u and r_v respectively, so the overall density power function can be defined as $\psi = \psi_u \cdot \psi_v$. The maximum overall density power $\psi_{\max}$ is found through the robust mean shift regression, and its corresponding fitted affine motion parameters are regarded as the background motion model induced from the moving camera.

3.2 Simple Differencing and Blob extraction

This algorithm first simply differences a pair of consecutive camera-motion-compensated frames in a video sequence, and then adaptively thresholds the difference image to obtain a binary image, which only consists of the information of moving targets with background removed. A rigid threshold is not sufficient to cope with problems of illumination variation such as glints, shadows, which is very common in real image sequences.

An adaptive thresholding approach to the difference image is necessary for effective detection of the moving target. In our algorithm, the mean of the local intensity distribution is simply used as local thresholding levels. This technique achieves effective segmentation of moving targets from the static background ground in our UAV video surveillance test. The blob extraction algorithm [Synder and Cowart 1983] is then applied to the binary image for extracting the detective moving targets with different labels. Finally, the essential information of each target blob such as the centre location, size and shape of the target is obtained.

4. EXPERIMENTAL RESULTS

We have developed a standalone C++ software package for moving target detection and target speed measurement from video sequences. Three examples of the experimental results for moving target detection and speed measurement are presented in Figure 2, Figure 3 and Figure 4 respectively.

Figure 2(a) and 2(b) are two consecutive frames of an aerial video sequence with oblique view. Figure 2(c) shows the selected feature points for tracking, and the corresponding translation feature points are detected in Figure 2(d) through the KLT sparse optical flow method. Figure 2(e) is the registered image through our robust camera motion compensation technique. Figure 2(f) shows that the moving car are detected and marked as red, which indicates that the location and the size of the target blob is obtained. The small windows (g) and (h) with the detected moving target are extracted from Figure (a) and (e) respectively. The compound phase correlation technique is then applied to measure the translational motion of the car between the small window Figure 2(g) and (h). Figure 2(i) is the registered image from (h) to (g) by the shift of the moving car. The estimated horizontal and vertical velocities of the moving car between the two consecutive frames Figure (a) and (b) are v_x =-7.012 and v_y =3.548 respectively. This is well approved by the cancellation of the car feature in Figure 2(j) that is the difference between the small window images Figure 2 (g) and (i).

Similarly, Figure 3(a) and 3(b) are two consecutive frames of a video sequence, in which three cars were tracked by a moving camera on a helicopter. Figure 3(c) shows the selected feature points for tracking, and the corresponding translation feature points are detected in Figure 3(d). Figure 3(e) is the adaptive thresholding difference between the reference frame and the corresponding registered image through our robust camera motion compensation technique. Figure 3(f) shows that the three moving cars are detected and marked as red, green and blue respectively. The corresponding locations and speed estimates of the three moving cars are shown in Table 2.

	Location	Speed (pixel)
Red target	x=119.11, y=259.06	v_x =7.25, v_y =-0.96
Green target	x=134.92, y=110.86	v_x =-7.82, v_y =1.55
Blue target	x=361.94, y=227.36	v_x =7.78, v_y =-0.92

Table 3. Location and speed estimates of the three detected targets.

Figure 4 presents an example of moving target detection and motion estimation, and its application in the target resolution enhancement from a thermal image sequence of a Land Rover travelling on road. The camera is fixed at a position but panning to track the fast moving vehicle. In this case, the background is moving while the vehicle is more or less in the same position within the image frame shifting in a very limited range. Figure 4(a) shows the frame 1 and Figure 4(b) shows the frame 30 of the image sequence respectively. The moving Land Rover shown in Figure 4(c) was successfully detected and extracted between the consecutive frames of the image sequence using our phase correlation image analysis system (PCIAS) software package. According to EPRC analysis between frames (from frame 0 to frame 30) in different time interval, the random shift of the extracted vehicle sub-scene is from less than $1/10^{th}$ pixel to two pixels, which is sufficient for super resolution image reconstruction (SRR). Figure 4(d) shows the SRR images of the moving Land Rover reconstructed from 30 consecutive scenes, which shows well-improved target resolution with the sharper edges. This SSR example test indicates that the proposed ERPC based target motion estimation technique is able to be applied in target resolution enhancement for video surveillance.

5. CONCLUSIONS

The analyses and experimental results presented in this paper have demonstrated that our moving target detection and motion estimation scheme, which combines the robust camera motion compensation, simple differencing with adaptive threshold and ERPC technique, is able to achieve sub-pixel accuracy for moving target speed measurement in aerial video sequence. In particular, the proposed ERPC method is capable of measuring the motion of a moving target in a stationary background as long as the target is the dominant feature in the image frame. In addition, the moving target speed measurement technique presented here can be applied in target resolution enhancement for video surveillance.

6. ACKNOWLEDGEMENT

The work reported in this paper was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence.

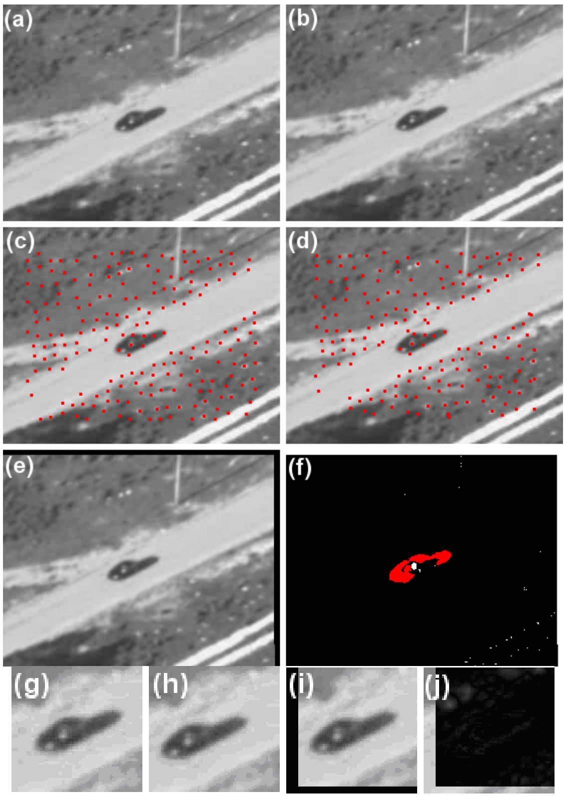

Figure 2: Moving target detection and speed measurement from two consecutive frames of an aerial video sequence with one moving car.

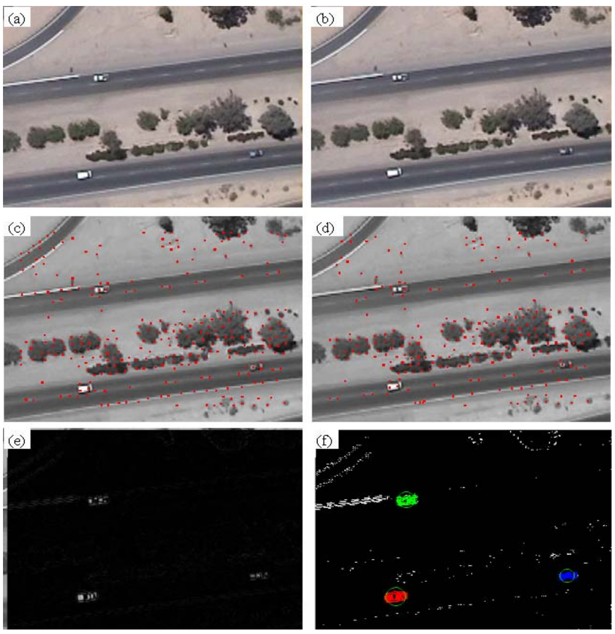

Figure 3: Moving target detection and speed measurement from two consecutive frames of an aerial video sequence with three moving car.

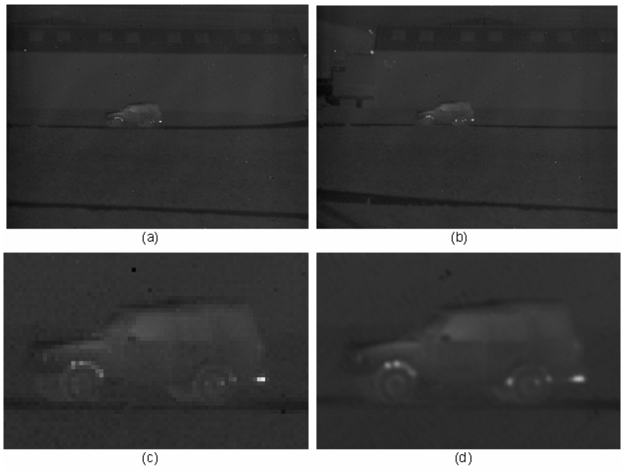

Figure 4: SRR of a thermal video image sequence of a moving vehicle. (a-b) Frame 1 and frame 30 extracted from the video with 314×233 pixels. (c) The image (76×49 pixels) of the moving vehicle extracted from frame 1. (d) The SRR result produced from 30 frames.

7. REFERENCES

Comaniciu, D. and Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*. vol. 24, no. 5, pp. 603-619.

Foroosh, H., Zerubia J. B., and Berthod, M., 2002. Extension of phase correlation to subpixel registration. *IEEE Trans. Image Processing*, vol. 11, no. 3, pp. 188-200.

Hoge, H. S., 2003. Subspace identification extension to the phase correlation method, *IEEE Trans. Medical Imaging*, vol. 22, no. 2, pp. 277-280.

Kuglin, C. D. and Hines, D. C., 1975. The phase correlation image alignment method. *Proceeding of IEEE International Conference on Cybernetics and Society*, pp. 163-165, New York, NY, USA.

Kumar, R., Sawhney, H., Samarasekera, S., Hsu, S., Tao, H., Guo, Y., and Hanna, K., Pope, A., Wildes, R., Hirvonen, D., Hansen, M., and Burt, P., 2001. Aerial video surveillance and exploitation, *Proc. IEEE*, vol. 89, pp. 1518–1539.

Liu, J. G. and Yan, H., 2006. Robust phase correlation methods for sub-pixel feature matching. *Proceeding of 1st Annual Conference of Systems Engineering for Autonomous Systems, Defence Technology Centre*, A13, Edinburgh, UK.

Shi, J., and Tomasi, C., 1994. Good Features to Track, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593–600.

Stone, H. S., Orchard, M. T., Chang, E.-C., and Martucci, S. A., 2001. A fast direct Fourier-based algorithm for subpixel registration of image, *IEEE Transactions on Geoscience and Remote Sensing,* vol. 39, no. 10, pp. 2235-2243.

Synder, W. and Cowart A., 1983. An iterative approach to region growing using associative memories, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wang, H. and Suter, D., 2004. A. A very robust estimator for modelling fitting and range image segmentation. *International Journal and Computer Vision*, vol.59, no. 2, pp. 139-166.

Wang, H. and Suter, D., 2004. B. Robust adaptive-scale parametric model estimation for computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 1459–1474.