

Temporal Resolution Enhancement from Motion

M. J. A. Strens, M. P. Rollason
QinetiQ Limited, 1105/A7, Cody Technology Park,
Ively Road, Farnborough, GU14 0LX

Abstract

We describe progress in the first year of the EMRS DTC TEP theme project entitled “Temporal Resolution Enhancement from Motion”. The aim is to develop algorithms that combine evidence over time from a sequence of images in order to improve spatial resolution and reduce artefacts. A C++ implementation of an initial algorithm has been developed, allowing assessment of the new algorithm against a number of datasets. Results are presented for resolution enhancement of static and moving scene content. The technique is shown to improve clutter rejection in target acquisition and to achieve useful resolution enhancement in maritime surveillance imagery.

Keywords: Image processing, super resolution, multi-frame methods.

Introduction

We set out to develop temporal resolution enhancement (TRE) techniques that exploit a sequence of images in which there is relative motion between sensor and scene/target, to provide step-change improvements in target acquisition, target identification and scene reconstruction performance.

An information theory argument suggests a benefit from processing multiple images. As we increase the number (m) of images processed, the amount of information available about the scene increases proportionately (given sufficient motion). In contrast, the dimensionality of the ‘state’ to be inferred is almost constant, because it is dominated by the scene description (rather than geometric and photometric transformation parameters that are relatively low-dimensional). Asymptotic analysis implies that spatial resolution will improve by up to \sqrt{m} in each axis compared with a single image frame.

The limitation in exploiting this information is in the ability to formulate and efficiently solve the inference problem. Significant progress has been made in this direction through the development of high-dimensional Bayesian inference approaches [8,10] and research in the image processing related disciplines of super resolution [1], optical flow [4], track before detect [9], structure from motion [3] and scene reconstruction.

Case studies using image data from a variety of military application domains will provide quantitative results and immediate push-through into higher technology readiness (TRL4-6) programmes for a range of airborne imaging systems (missile, UAV or fast jet). In this paper, we summarise results for TRE of a moving target in IR maritime surveillance imagery, and for clutter rejection in long-range IR target acquisition.

Military Relevance

TRE will improve the effective resolution of legacy hardware through software-only

upgrade. For future systems it will enable the use of smaller, lighter and cheaper sensors for a given required level of performance. We now detail three application areas for the technology.

1. TRE will obtain enhanced resolution **target images** for presentation to a human or to an automatic target identification system, leading to increased identification range in EO/IR imagery, active and passive. There is a strong demand for the technologies in fast-jet air-to-surface systems. There is also potential for immediate exploitation into airborne maritime surveillance systems, where target TRE could improve human identification of fast inshore attack craft (FIAC) and similar threats.

2. TRE will provide enhanced resolution (background) **scene images** in a moving-sensor system (weapon, UAV, conventional aircraft, ship, land vehicle). This will multiplicatively improve target detection performance in modern systems that detect targets as outliers from the background (e.g. using their relative motion), especially in low pixel-count legacy sensors. Technology insertion into low-cost software-only upgrades of air-to-air missiles is feasible.

3. TRE will provide enhanced resolution scene reconstructions or **mosaic images** for mapping applications and automatic enhancement of surveillance and reconnaissance imagery. An example is UAV mapping of an urban area prior in preparation for a military mission.

Technical approach.

Consider the two dimensional (2D) inference problem. For a Bayesian formulation, we require a *generative model* of the sensor system. This models the process by which the low resolution image I measured at the detector array is formed from the actual scene S . The series of

transformations (**Figure 1**) is conditional upon some photometric parameters P (e.g. gain and offset of the sensor) and geometric parameters A (e.g. the aspect of the sensor to the scene). The model must account for atmospheric attenuation, the optical system, and detector performance. **Figure 2b** shows the effect of optical blur and detector aliasing on the scene in **Figure 2a**. Adding detector noise yields the observed signal (I) shown in **Figure 2c**.

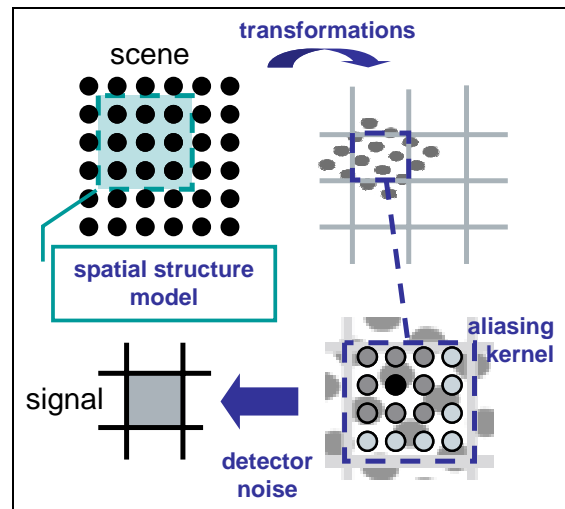


Figure 1 The generative model.

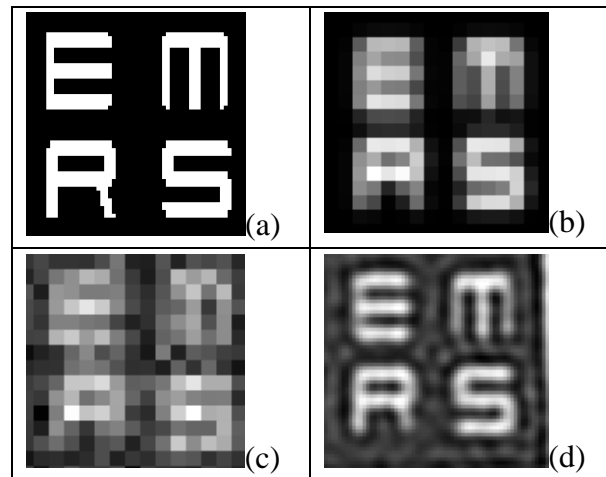


Figure 2 (a) Scene; (b) optically blurred; (c) noise added; (d) TRE result.

The generative model can be written $P(I | S, P, A)$. The inference problem is to obtain $P(S, P_{t_r}, A_{t_r} | I_{t_r})$ (where time-indexing has been introduced to represent a batch of data). Other information that must

be made available includes prior distributions on S , P and A . The prior distribution on S will typically encode spatial structure constraints. (These have had extensive research because they are the basis of single frame super resolution and image compression.) In order to process a batch of images, any temporal dependency between successive realisations of P and A must be expressed in the form of a dynamics model. For example, the dynamics $P(A_t | A_{t-1})$ on the geometric parameters expresses the rate of change of viewing aspect; *i.e.* the optical flow.

This is a challenging inference problem in Bayesian terms: a target tracker or navigation estimator would typically have 5 or 10 state variables whereas our super-resolved scene has thousands (one per sample point). Therefore decomposition of the inference problem into smaller parts and further approximations are required. The challenge is to ensure that the decomposition does not introduce significant loss of accuracy compared with the full formulation.

Significant experimentation with different choices for the decomposition led us to a solution that relies on ‘single point’ estimates for the scene (rather than multiple hypotheses), but within an iterative framework that ensures early (‘bootstrapping’) errors are gradually eliminated. If full posterior densities were to be introduced in place of maximum likelihood estimates, our experiments suggest that it is the low-dimensional parts of the state space (the photometric and geometric parameters) that will benefit most.

Algorithm Overview

The most important quantities are the “Scene” estimate (which is stored as a grid of points) and a list of “View” structures that each contain the geometric and

photometric transform parameters for a single frame.

Each incoming frame (sensor image) is considered in the context of the current scene estimate, but also with reference to a sliding time window of the last n frames, where typically $n=25$. **Figure 3** illustrates the processing of each new frame (at time t), which proceeds in 2 stages.

The first stage is a multi-scale search in the geometric parameters, to bring the incoming image into rough alignment (typically $\sim 1/4$ pixel accuracy) with the stored scene estimate. Within this process, maximum likelihood values for the photometric parameters (gain and offset) are obtained directly. On subsequent iterations the same image is revisited and its alignment improved, until it moves out of the time history.

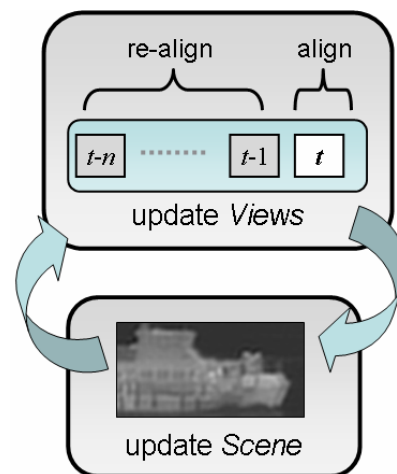


Figure 3 Algorithm overview.

The second processing stage updates the scene estimate to take account of the new image. It iterates through all scene points (in pseudo random order) and updates their values based on a conditional distribution relative to their neighbours. This conditional distribution is derived from both the generative model (applied over every low resolution pixel to which the scene point contributes during the sliding time window) and from the spatial structure model. This process is very different from

temporal averaging because it directly exploits the generative model. It is essentially a decomposed (local) version of the Landweber iteration commonly used in linear inverse problems and specifically in image deconvolution.

Details of scene point update

More formally, consider the update to scene point x with intensity $S(x)$ in order to reduce the reconstruction error over a batch of image frames (the sliding window). Within each of these images, scene point x may contribute to a number of low resolution cells (pixels), depending on the point spread function of the optical system. The error at each of these low resolution cells is the difference between the observed pixel intensity and its predicted (reconstructed) intensity under the generative model. The latter quantity is obtained by application of the generative model to S using the current alignment estimate. The cost function of interest is the sum of squares of these errors over the batch of images. The differentiability of the generative model allows us to compute how this cost will be affected by a change in $S(x)$.

This error criterion is not sufficient to obtain an effective update to $S(x)$ because the system is underdetermined: some ‘regularisation’ penalty must be applied to avoid overfitting (which leads to microstructure in the scene estimate). A *spatial structure model* achieves this effect by penalising large gradients (or other measures of variability) in S . In the Bayesian view, the spatial structure model is equivalent to a prior probability density on S . We have explored several such models, but the experiments in this paper make use of a simple second derivative constraint. Specifically, the penalty is proportional to the square of the difference between a scene point and the mean of its four immediate neighbours.

The gradient of this penalty is known and can be combined with the gradient of the reconstruction cost function to obtain a suitable change in $S(x)$. The magnitude of the change can be chosen so as to minimise the total cost, or can be scaled by some learning rate chosen to maximise the rate of convergence for the scene as a whole. Asymptotic convergence is guaranteed for any learning rate below unity.

Experiments with Synthetic Data

Figure 2d shows the result of applying the algorithm to 100 frames of a synthetic dataset, in which there was a random spatial jitter (r.m.s. 1 pixel) in each frame, and additive detector noise (SNR = 10). Spatial structure that was aliased in the individual observations (**2c**) was recovered in the scene estimate. Experiments with further synthetic test patterns verified the theoretical prediction that image resolution in each axis will improve in proportion to the square root of the number of frames used, *if* there is sufficient structure in the scene to achieve accurate alignment.

However, experiments with synthetic data can be misleading, because they do not account for ‘modelling error’: the disparity between the assumed generative model and the way the data was actually formed. This can be catastrophic for Bayesian methods, and therefore evaluation with real image datasets are a major focus of this project.

Software Description

The TRE process is computationally demanding, and so it was determined at an early stage that an efficient implementation (in a compiled language) would be needed to perform evaluation with significant amounts of test data. Therefore a C++ implementation of the algorithm has been developed and refined during the first year of the project. Object-oriented design features of the C++ language have been

exploited to obtain a flexible and extensible implementation to support both algorithm research and large-scale evaluation. Parameters defining the input imagery, the generative model, and the processing to be applied are specified through a settings file, allowing easy configuration of experimental evaluations.

The key ‘inner loop’ processes are (i) the calculation of conditional densities in the scene update, and (ii) the reprojection of the scene for computing residuals in the alignment process. For the affine family of geometric transformations, process (ii) is an image convolution (for which many hardware acceleration options are available), but process (i) remains challenging and a different formulation of the inverse problem may be required to achieve real time performance.

Experiments with maritime IR target

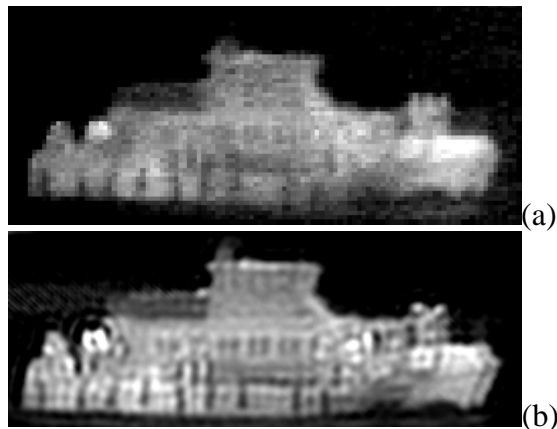


Figure 4 (a) *Input data (ferry)*; (b) *TRE result*.

We have performed an evaluation of the TRE processing using imagery collected with a Wescam MX15 imager under the DEC AWE AIMMS programme. Successful resolution enhancement has been obtained for a large ferry at constant velocity, and for a smaller boat that has significant rotational motion. The ferry sequence has been used to perform comparative evaluation of processing options. In **Figure 4** there are several details (such as the shape of the hull) that

can be seen in the TRE result but are not apparent in the raw image.

To quantify the benefit, the standard measure is the root mean square (RMS) reconstruction error, which measures the residual difference between the input data and the model for that data. **Figure 5** shows that the RMS error is significantly reduced by TRE compared with a baseline process that accurately aligns the images but does not attempt super resolution. The error is minimised when the frame is in the middle of the (16 frame) sliding window. It should be noted that the RMS error can go no lower than the detector noise level, which is estimated at ~ 0.03 for this sequence. **Figure 6** shows the benefit of the realignment stage in our processing architecture. Clearly there is a major benefit from reinterpreting the sliding window history in the context of the current scene estimate.

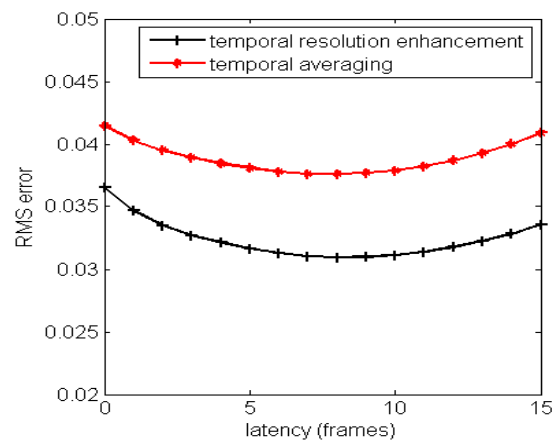


Figure 5 *Comparison with averaging.*

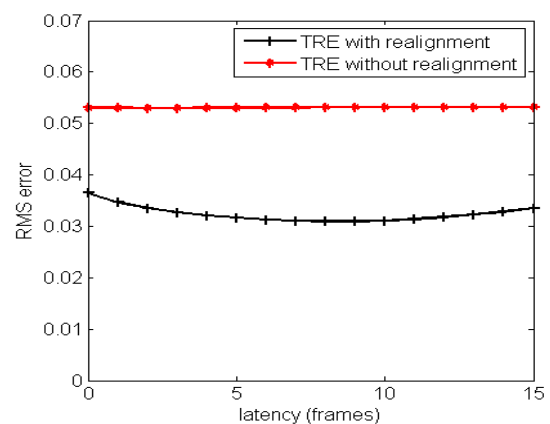


Figure 6 *The benefits of realignment.*

Clutter rejection in acquisition

Temporal resolution enhancement may be of significant benefit in the acquisition of difficult targets. Modern target detection systems are generally limited in their performance by clutter in the scene, especially clutter to which a target filter has a strong response. However, if there is relative motion between the target and the scene background, it should be possible to obtain an accurate estimate of persistent clutter and hence eliminate it.

In previous research for the DEC TA Weapon Software Insertion (WSI) programme we have shown that obtaining a scene estimate can significantly reduce clutter and increase acquisition range. When working with a low pixel-count legacy sensor, we found that some artefacts remained because the scene estimate was not adequately (spatially) sampled. Therefore we now evaluate whether building a super resolved scene estimate can further improve acquisition range, using image data supplied (with IPT permission) by the WSI programme.

We applied the TRE processing to a sequence of ground-based data collected with an IR missile seeker head, containing significant levels of clutter (cloud and treeline). We compared the Receiver Operating Characteristic (ROC) for three cases: (i) single frame target acquisition (no clutter suppression); (ii) multi-frame clutter map at seeker resolution; (iii) super-resolved clutter map from TRE processing.

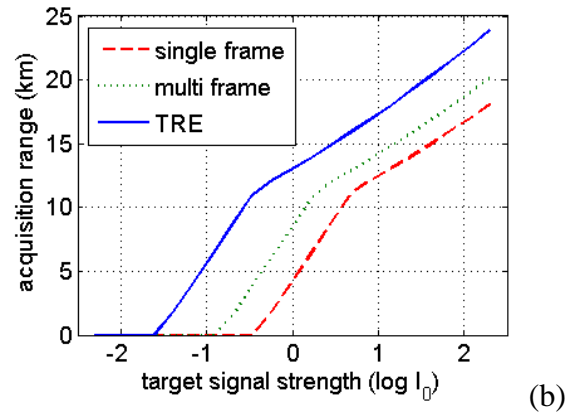
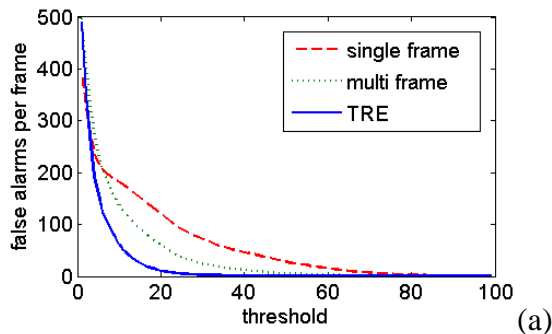


Figure 7 Acquisition performance comparison: (a) ROC curves; (b) range estimate.

From **Figure 7a** we observe that TRE has a major clutter suppression effect, reducing the number of false alarms that will be passed from the acquisition image processing to the next stage. **Figure 7b** indicates what the impact of this processing would be on acquisition range. This is obtained assuming a constant false alarm rate (CFAR) and using a physical model that relates target signal strength to range (taking account of atmospheric attenuation and sub-pixel effects). We see that TRE offers a *potential* acquisition range improvement of between 5 and 10 km at “typical” target signal strengths (-1 to 1 on the axis). However, this result was obtained with only a single sequence of data and a much larger assessment would be required to quantify the benefits for whole system performance.

Related Work

TRE is multi-disciplinary research because it combines a range of image processing topics with the need to solve a major computational problem that may be expressed in the Bayesian framework (as an inference problem) or as cost function minimisation (linear or nonlinear inverse problem).

Pickup et al. [7] recently proposed a Bayesian multi-frame super-resolution technique that uses a similar generative

model to our own. Their approach was not evaluated with real video sequences (rather than sequences synthesised from the generative model) and so it is not yet clear how robust it will be with respect to modelling error. Nevertheless they offer an alternative scene update method (“scaled conjugate gradient”) and a different spatial structure model (nonlinear, first-derivative based).

Farsiu *et al* [2] propose a nonlinear method based on an L1 norm (i.e. absolute difference) criterion for both parts of the error function. They claim that this improves the sharpness of edges in the scene estimate by rejecting outliers. We have implemented the L1 based “total variation” spatial structure model and found no significant benefit to compensate for the increased computational cost of moving to a nonlinear model, but we will investigate the application of the L1 criterion for reconstruction error because this could have a major benefit in long sequences (in the presence of modelling error).

In order to compare with such approaches it is our aim to implement a library of alternatives for each of the processing stages. This is already true of the spatial structure model where we have compared the simple linear model used herein with the total variation method and a competitive multi-model approach.

Conclusions and Future Work

We have developed a flexible algorithm for temporal resolution enhancement that has broad applicability for improving the performance of imaging sensor systems for a variety of tasks including acquisition and target identification. We have focused on the need for an efficient decomposition of the inference problem that reduces computational costs but retains accuracy, and on addressing the modelling error that

is inevitable when working with genuine sensor data.

The algorithm has been shown to be effective for resolution enhancement of slow-moving maritime targets (and static scene content). Motion blur is not yet accounted for; including integration over time in the detector model is a straightforward and necessary improvement to achieve enhancement of fast-moving targets that are not being tracked closed-loop.

The emphasis in the second year of this research will be on the resolution enhancement of airborne surveillance imagery. This type of data should provide good alignment accuracy and therefore major benefits are expected, but there are computational challenges (in working with much larger quantities of image data), and a need to account for distortion due to nonlinearities and depth variation in the scene geometry. The latter requirement suggests the need for a scene model that can: (i) account for out of plane rotations of target or scene; or (ii) augment the scene with depth information.

As we move to working with larger datasets and more sophisticated scene representations, it is expected that more advanced inference methods will need to be applied in order to reduce computational load. For the scene update stage, the current approach calculates the error gradient for individual scene points and updates them one at a time. However, it is likely that this process is relatively inefficient compared with an update to a whole block of nearby scene points (re-using some of the generative model calculations in the process).

For the alignment stage, it is likely that the current parameter space search approach will become less efficient as the dimensionality of the (geometric and

photometric) transformations is increased. Furthermore, there may be a benefit in obtaining a full Bayesian posterior for these parameters. The HINTS algorithm [10] is ideally suited to this problem because it can exploit the additive structure of the error function to obtain a very rapid search of the multi-dimensional parameter space.

There is also potential for a broader set of exploitation routes for the research, including application to active (*e.g.* burst illumination laser and lidar) sensors in EO/IR, and to other wavebands. While the imaging geometry of active and coherent sensors can be different, a similar large scale inference problem needs to be solved.

References

1. Baker S and Kanade T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24:1167-1183, 2002.
2. Farsiu S *et al.* Fast and robust multi frame super resolution. *IEEE Transactions on Image Processing*. 13(10):1327-44, 2004.
3. Faugeras O. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.
4. Irani M and Peleg S. Super Resolution From Image Sequences, *ICPR*, 2:115--120, June 1990.
5. Landweber L. An iterative formula for Fredholm integral equations of the first kind. *American Journal of Mathematics* 73:615–624, 1951.
6. Lucas B D and Kanade T. An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging understanding workshop*, pp 121—130, 1981.
7. Pickup L C *et al.* Bayesian Image Super-Resolution, Continued. *Advances in Neural Information Processing Systems*, December 2006.
8. Strens M J A *et al.*. Markov Chain Monte Carlo Sampling using Direct Search Optimization, *Proceedings of the Nineteenth International Conference on Machine Learning*, San Francisco: Morgan Kaufmann, 2002.
9. Strens M J A and Gregory I N. Tracking in Cluttered Images, *Journal of Image & Vision Computing* 21(10), pp 891-911, Elsevier, 2003.
10. Strens M J A *et al.* Efficient Hierarchical MCMC for Policy Search, accepted for the Twenty-first International Conference on Machine Learning, 2004.

Acknowledgements

The work reported in this paper was funded by the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by a consortium SELEX Sensors and Airborne Systems, Thales Defence, Roke Manor Research and Filtronic.