

Temporal resolution enhancement from motion - application to airborne imagery

Malcolm Rollason, Graham Watson
QinetiQ Limited, 1105/A7, Cody Technology Park,
Ively Road, Farnborough, GU14 0LX

Abstract

We describe progress in the third 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 unwanted artefacts. Years one and two of this project developed and demonstrated an efficient algorithm that provided good resolution enhancement of a scene viewed in the far field (approximately flat) [1]. This paper reports a new algorithm which is applicable to a three dimensional scene where substantial depth variation causes parallax within the imagery. The new algorithm is demonstrated using airborne infra-red imagery.

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. 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 and research in the image processing related disciplines of super resolution, optical flow, track before detect, structure from motion and scene reconstruction (see the references contained in [1]). 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 (surveillance EO turret, missile, UAV or fast jet).

Good resolution enhancement has been achieved for a scene viewed in the far field (approximately flat) as reported in [1]. This paper describes a new algorithm for resolution enhancement of a scene containing substantial depth variation.

We demonstrate TRE of infra-red airborne imagery to produce resolution enhanced estimates of (i) an (approximately flat) scene viewed in the far field (ii) a three dimensional scene where substantial depth variation causes parallax within the imagery.

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 two 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 both EO and IR imagery. There is a strong military need for the technologies in fast-jet air-to-surface systems. There is also potential for immediate exploitation into airborne maritime surveillance systems, where TRE could improve human identification of fast inshore attack craft (FIAC) and similar threats.

2. TRE will provide enhanced resolution scene (background) images in a moving-sensor system (weapon, UAV, conventional aircraft, ship, land vehicle) to aid clutter suppression in target acquisition.

Technical approach

Our approach to TRE for an approximately flat scene which is viewed at a distance was described in [1].

For both 2 D and 3 D scene content we have formulated TRE as a Bayesian inference problem where the unknown state variables are the super resolved scene and the transforms between successive frames. To solve the inference problem requires careful modelling of the process by which the image was formed. Hence a generative model is used in the inference process to obtain estimates for the unknowns, given the observed sequences of data. We have developed an efficient framework for solving this inference problem by iterating between an efficient Bayesian alignment process that estimates the geometric and photometric parameters, and a scene update stage that re-computes the super-resolved image (**figure 1**).

The choice to decompose the inference in this fashion allows application of methods appropriate to the dimensionality of the problem :

- Scene points $\sim 10^5$ (a grid of intensities supersampled with respect to the sensor image)

- Transformations ~ 10 times the number of frames

The iterative inference scheme is depicted in **figure 1**. On receipt of a new image, the first stage is identification of the geometric parameters, to bring the incoming image into **alignment** with the stored scene estimate.

The second stage is an **information update** where the new sensor image is used to update the scene estimate.

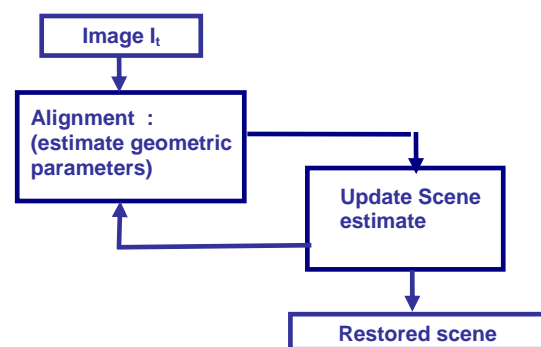


Figure 1 : Algorithm structure

Alignment

There are two key components of the alignment problem :

- the scheme used to identify the alignment parameters in the geometric model;
- the geometric model used to map scene points to image points.

An efficient Bayesian inference scheme (HINTS) is used to infer the geometric transformations. In year 2, the HINTS scheme was included within the framework and used to infer the full affine parameter set. In year 3, HINTS has been used to infer parameters within a rigid body transform which accounts for depth variation within the scene.

The HINTS scheme is ideally suited to the alignment problem because it can exploit the additive structure of the error function to obtain a very rapid search of the multi-dimensional parameter space. The scheme

is detailed in [2] and uses a hierarchical structure (**figure 2**) where changes in state at the top (root) level are the result of sequences of Metropolis-Hastings steps taken at the lower levels.

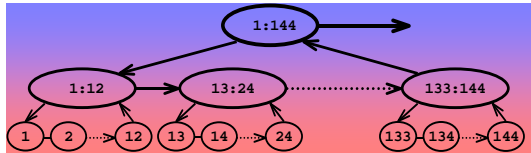


Figure 2 : *HINTS sampling architecture*

An important part of the framework is the geometric model which accounts for the motion between frames in incoming imagery. The motion must be modelled correctly since it is used to provide the resolution enhancement. We have used two different geometric models within the framework :

- (i) an affine transform - applicable to an (approximately flat) scene viewed in the far field (2D algorithm);
- (ii) a rigid body transform – applicable to a three dimensional scene where substantial depth variation causes parallax within the imagery (3D algorithm).

Whichever model is used, it performs the function of mapping scene points to image points and hence describes the alignment of an incoming image relative to the stored scene estimate.

To predict a sensor resolution image from the scene estimate, the geometric model is first applied to project the scene into the co-ordinates of the inbound image. The generative model is then applied to model the optical effects and produce a predicted sensor image.

An affine geometric model was discussed in [1]. It uses image based transformations and is effective only for scene content that is planar and viewed at range.

This year we extended the framework to allow for a single rigid body motion with

arbitrary depth discontinuities. This requires a three dimensional representation of the scene, and a three dimensional transform. A rigid body motion consists of a 3D displacement and a 3D rotation and thus the motion between frames is described by six parameters. As with the affine transform, these parameters are inferred using the HINTS scheme.

Information update

The second processing stage (**figure 1**) updates the scene estimate to take account of the new image.

The differentiability of the generative model allows an error gradient to be calculated for individual scene points. An efficient *whole scene* update scheme was developed in year 2. This *gradient based* scheme seeks to minimise an error function comprising (i) a local smoothness penalty and (ii) a reconstruction error computed as the difference between measurement image pixel value and the corresponding pixel value predicted via the generative model, given the estimated transformations and scene. This type of gradient update scheme is a Landweber iteration and is very different from temporal averaging because it directly exploits the generative model.

The *whole scene* gradient update scheme has been implemented using whole image operations (ie convolutions) so that the update scheme could be carried out quickly by parallel architectures such as a FPGA.

Algorithm summary

Key elements of the TRE processing can be summarised as follows. The TRE algorithm :

- models spatial structure (local smoothness) in the scene;
- models the sensor point spread function and detector noise;
- combines a generative model of the image formation process with the estimate of the resolution enhanced

scene and transformations to predict sensor resolution measurements;

- updates the scene estimate so as to minimise the error between each received image and the image predicted from the scene estimate;
- infers full geometric and photometric transformations

Results

Airborne IIR imagery sequences from an MX series turret have been made available by L-3 WESCAM. This section presents three results for types of scene :

- *an approximately flat scene viewed in the far field;*
- *a three dimensional scene where substantial depth variation causes parallax within the imagery .*

Resolution enhancement of an approximately flat scene

A region of interest (ROI) identified initially by an operator is tracked by the algorithm as it moves within the airborne sensor's field of view (FOV).

Figure 3 shows an example sensor image and the TRE output after 100 frames. It should be noted however that most of the resolution enhancement has been achieved after ~30 frames.

This result shows good resolution enhancement and clarifies details such as the number of spokes in the car's front wheel that are not visible in the sensor imagery even when viewed as a sequence.

A further example which illustrates resolution enhancement of an approximately flat scene viewed at an oblique angle is provided in **figure 4**.

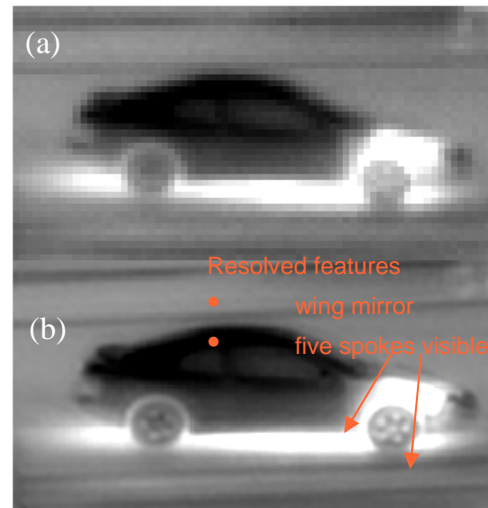


Figure 3 : (a) IR sensor image (41 by 71 pixels); (b) TRE result for a target tracked as it moves within the sensor FOV.

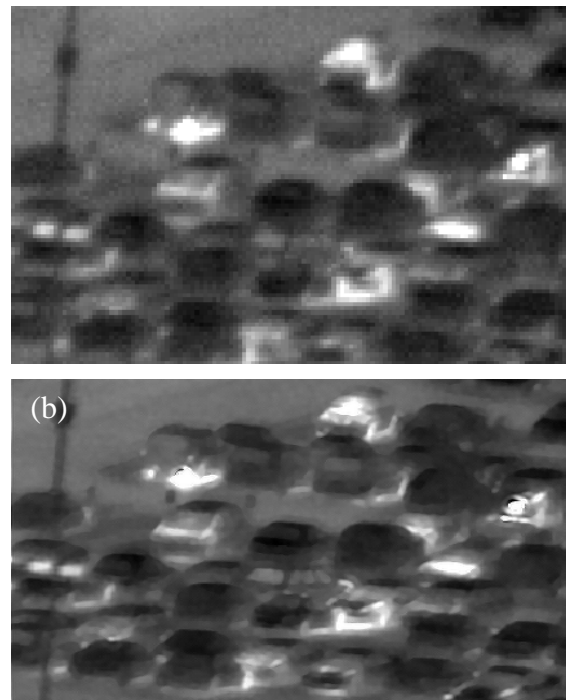


Figure 4 : Resolution enhancement of an obliquely viewed scene; (a) Sensor image; (b) TRE result

Resolution enhancement of a three dimensional scene

The main algorithm development in year three has enabled, for the first time, resolution enhancement of a 3 D scene with strong parallax. Three dimensional structure in the scene is inferred from the passive imagery.

The new 3D algorithm :

- provides resolution enhancement of a scene with strong parallax;
- uses a new scheme for estimation of the dense depth map to provide a stable scene estimate.
- Uses the efficient Bayesian scheme (HINTS) to estimate Rigid Body Transform coefficients.

Two frames in the input image sequence (**figure 5**) show strong depth dependent motion of objects in the scene. There is strong parallax evident – note that the car and sign posts move right to left behind the right hand telegraph pole, which similarly moves right to left behind the central pole which is in the ‘middle distance’.

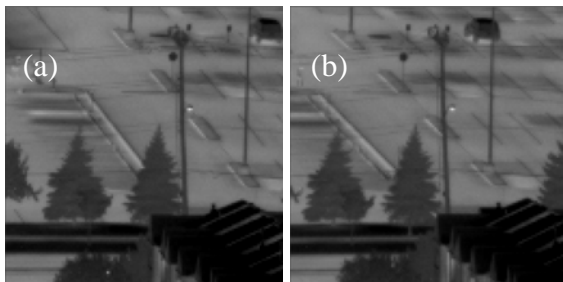


Figure 5 : images in input sequence, 3D scene content: (a) frame 10; (b) frame 83.

Figure 5 shows two frames in the input image sequence show strong depth dependent motion of objects in the scene. In this scene there is strong parallax - the car and sign posts move right to left behind the right hand telegraph pole, which similarly moves right to left behind the central pole which is in the ‘middle distance’.

The 3D TRE processing result is shown in **figure 6**. The estimate of scene intensity is shown on the left and shows improved resolution of car and telegraph pole. The depth map (range from sensor to scene) is shown in **figure 6** (b) where objects estimated to be at greater range appear lighter in colour. The car and right hand pole are correctly estimated to be further from the sensor than the central telegraph

pole, which is in the ‘middle distance’ and hence is neither black nor white in the depth map. An expanded view of the results is shown in **figure 7**.

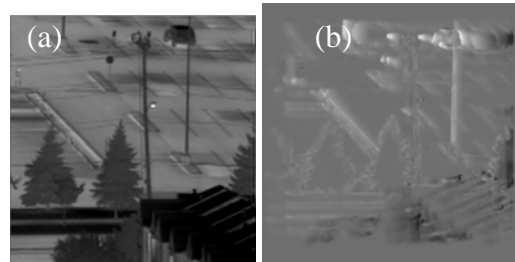


Figure 6 : TRE output showing scene estimate (a) intensity; (b) dense depth map.

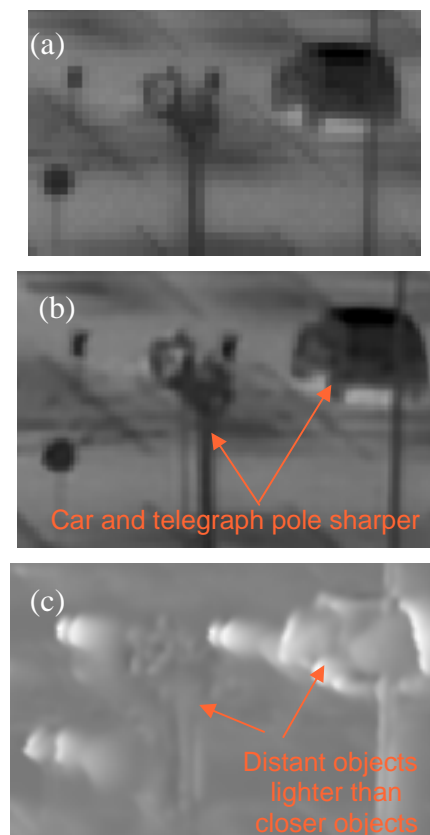


Figure 7 : expanded view (a) input image; (b) TRE intensity estimate; (c) TRE depth estimate.

Quantified improvement in resolution

We have developed a metric to quantify the resolution enhancement provided by our technique which identifies the size of the smallest object that we would expect to be clearly resolved in the TRE result, and for

comparison, also in a Temporal Average (T.A.) result, and deconvolved T.A. image.

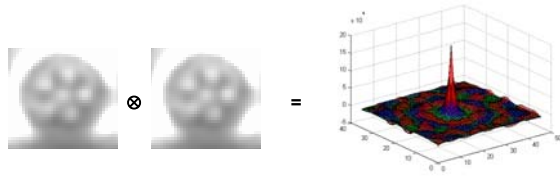





Figure 8 : *autocorrelation surface*

The metric is based on an autocorrelation surface computed for a detailed part of the scene, **figure 8**. The width of the central lobe is parameterised, and the equivalent point spread function identified - leading to the results quoted in **Table 1**. Values in the table identify the smallest resolveable object size, in units of pixels in the super-sampled scene, and equate to twice the standard deviation of an equivalent point spread function (e-psf) for each image.

Sequence	e-psf (2σ)		TRE psf as % of T.A.
	T. A.	TRE	
	5.76	2.48	43 %
	6.00	2.47	41 %
	3.88	2.82	72 %

The last column identifies the width of the e-psf for the TRE output as a percentage of the corresponding width of the temporal average result. Values of less than 50% equate to more than a doubling of resolution – a factor of between 2 and 3 is achieved for the 2D algorithm applied to an approximately planar scene viewed at a distance. A lesser, although still significant improvement, is achieved with the 3D

algorithm for a sequence with strong depth discontinuities and parallax.

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.

Good resolution enhancement has been demonstrated for an approximately planar scene viewed in the far field.

A new algorithm has been developed this year for resolution enhancement of a scene containing substantial depth variation. This algorithm is valid for a single rigid body motion.

We have developed a metric to quantify the resolution enhancement produced by the algorithm. Resolution enhancement by a factor of more than 2 is achieved for 2D scene content, and a factor of ~1.4 for 3D scene content. Further work should improve the resolution gain, and efficiency, of the 3D algorithm. Further development should also include modelling of multiple object motion in the scene.

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 of SELEX Galileo, Thales UK and Roke Manor Research.

References

1. Rollason, M P and Gardner A P, "Temporal Resolution Enhancement from Motion – application to airborne imagery" Electro-Optic Systems, Embedded Processing and Devices,

EMRS DTC Annual Technical Conference, 2008.

2. Strens M J A et al.. Markov Chain Monte Carlo Sampling using Direct Search Optimization, Proceedings of the Nineteenth International Conference on