Automatic Sightline Stabilisation in Noisy Imagery

SHANGQI BAO, JASON F. RALPH
Department of Electrical Engineering & Electronics
University of Liverpool
Brownlow Hill, Liverpool, L69 3GJ.
UNITED KINGDOM
http://www.liv.ac.uk/EEE/

Abstract: - This paper examines the use of automatic image registration for the stabilisation of images from a moving camera in the presence of noise. In good quality imagery, the motion of the camera platform can be compensated by matching the features from image to image and deriving the appropriate image transformation that will correct for the motion of the camera. This paper considers the sensitivity of such registration algorithms to the different types of noise that occur in infrared imaging. The most reliable method for the registration of infrared imagery in the presence of these noise sources were Fourier registration and region-based registration. Both of them are robust and the Fourier method had the added advantage which is relatively insensitive to the initial image translation errors.

Key-Words: - Image Registration, Fourier Registration, Image Stabilisation, Noisy Images

1 Introduction

This paper examines the use of automatic image registration [1] for the stabilisation of images from a moving camera in the presence of noise. This represents a significant problem in infrared imaging from an unsteady or moving platform. In many cases, supplementary information is available from auxiliary sensors, such as gyroscopes or inertial measurement units, but these tend to be expensive. A simpler, but more computationally expensive approach is to use the ego motion of the camera – the ego motion is the camera motion that is derived from the apparent motion of the (static) background [2]. To generate reliable estimates of the camera motion and to calculate the appropriate geometric transformation that will remove the effect of the motion from the image, a high contrast stable scene is normally required. Unfortunately, infrared imagery is often subject to large amounts of fixed and variable pattern noise [3]. In particular, an infrared imager tends to have a far lower contrast in a normal scene than a conventional visible band imager and is somewhat sensitive to thermal glare from very intense hot spots, which tends to suppress the relative contrast in other areas of the image. In addition, many effects that are not visible in the visible band become apparent in the infrared bands (mid-wave infrared = 3-5 microns and long wave infrared = 8-12 microns wavelength). A relatively common example is the appearance of exhaust plumes from the engine of the aircraft, which fluctuate rapidly as the aircraft flies and can mask the features on the ground. This variable pattern noise can make it very difficult to correlate the background scene from image-to-image (see Fig.1 for an example) and can therefore prevent reliable image stabilisation. This is particularly important where the camera has a very small field of view and the effect of small movements of the platform will be magnified in the image plane.

Once the camera has been adequately stabilised, other processing can be performed, such as integrating successive images to enhanced image contrast to remove the adverse effects of the noise – noise removal.

This paper considers the standard techniques that are used for automatic image registration and compares their robustness for the different types of noise that could be present in imagery from an infrared camera mounted on a moving aircraft.

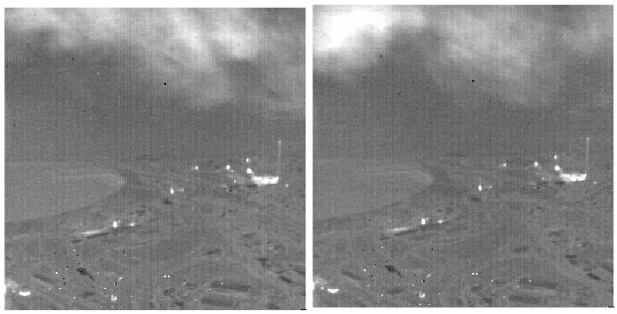

Fig.1 – Two successive frames of infrared imagery showing large changes in image noise from aircraft exhaust plume at the top of each image.

2 Image Registration

There are three main types of image registration algorithm that are in general use: feature based techniques, region/area matching techniques (local or global), and Fourier-based methods (a range of useful references is contained in [1]). The general properties of the three approaches are described in sections 2.1 – 2.3.

2.1 Feature-based methods

The simplest method of image registration is to extract features from each of the images and then work out the transformation by comparing the position of the identifiable feature points in one image with the position of the corresponding feature in the other image. This widely used technique is reliant on the ability of the algorithm to extract related features in each image and then associate the features found in one image with the features found in the other image. This can be computationally expensive when the number of features is relatively large.

The type of features to be detected will depend upon the type of imagery being matched (infrared, visible band, etc.) and the expected scene content. Detecting right-angled corners would be useful for city or urban environments but less useful for small-scale scene-matching applications. To deal with this variability, the algorithms investigated in this work contain a number of different feature detection techniques and are intended to detect different types of features.

Once the features have been detected in each of the images, one needs to find which of the features in one image corresponds to a feature in the other image. This is potentially the most time consuming process in feature-based registration, since a large number of features generates a large number of possible associations between features in each image. However, the number of features that need to be correctly detected and associated to generate the correct two-dimensional image transformation is relatively small – e.g. only two pairs of associated feature points are required to generate a two-dimensional affine transformation. As a result, it is often unnecessary to match all of the available features.

One of the simplest methods to associate features between images is to generate the Hu invariant moments [4] for the area around each of the detected features and then compare the moments to find the pairs that are 'closest' (in the Euclidean sense). The closest feature pairs are then stored and then each of the pairs is checked to see if they generate a valid image transformation.

2.2 Region-based methods

Region-based image registration methods differ from feature-based methods in that they use a large section or all of the images and match the whole region rather than simply associating selected points from one image to points in the other. There a two widely used measures to check the similarity of two images: the normalised cross-correlation [5] and the mutual information [6]. The approach of each method is to find the image transformation that maximises the similarity measure for the areas of the images that overlap once the transformation has been applied.

2.2.1 Maximum Cross Correlation

The (normalised) cross correlation between two images is given by [5]

$$Corr(f,h)=\frac{\sum_{x,y}(f(x,y)-\overline{f})\cdot(h(x,y)-\overline{h})}{\sqrt{\left(\sum_{x,y}(f(x,y)-\overline{f})\right)^2\cdot\left(\sum_{x,y}(h(x,y)-\overline{h})\right)^2}}$$

where f and h are the images, and $\overline{f}$ and $\overline{h}$ are the corresponding image mean values. The normalised cross correlation should be one when the images are exactly the same, minus one if they are negative images and some value between 1 and −1 if there are differences in the images. The closer the images are, in terms of content, the larger the cross correlation value will be. To find the transformation that maximises the value of the cross correlation, a number of search strategies may be employed. However since the motion of the camera from frame to frame should be relatively small, a gradient based-search strategy is employed here, although it does contain a stochastic element to avoid small local minima in the correlation function [6].

2.2.2 Maximum Mutual Information

The other similarity measure used is the mutual information for the joint probability distribution for the grey levels in each image, $p(a, b)$ [6]. The joint distribution is constructed by forming an $n\times n$ array, where n is the number of grey levels in the images. The elements of the probability array contain the number of pixels N_{ij} that have grey level i in image 1 and grey level j in image 2, divided by the total

number of pixels (assuming that the images are the same size). For images with a reasonable number of grey levels (8 bits = 256 grey levels) this array will be quite sparse – which gives a low value for the mutual information – so it is often necessary to reduce the number of grey levels by removing the least significant bits from the images, 64 grey levels or 6 bits is typical [6]. The mutual information is then calculated from,

$$I(B;A)=\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} p(a_i,b_j)\cdot\log_2\left(\frac{p(a_i,b_j)}{p(b_j)\cdot p(a_i)}\right)$$

where $p(a_i, b_j)$ is the joint probability distribution, and $p(a_i)$ and $p(b_j)$ are the single image grey level probability distributions (found by summing the joint distribution over the other image). As with the cross correlation, the greater the similarity between the two images the greater the value of the mutual information, so image matching can be done by finding the transformation that maximises the mutual information between the two images – using the same stochastic methods used for the cross-correlation calculations and described in reference 6.

2.3 Fourier methods

The Fourier-based techniques [7] use the fact that the translation of a signal (or other function) is represented by a phase shift in the Fourier domain. If two images are related by a translation

$$f(x,y)=g(x-x_0,y-y_0)$$

the Fourier transform of g will be related to the Fourier transform of f by a complex phase factor.

$$F(u,v)=G(u,v)\cdot\exp(2\pi j(ux_0+vy_0))$$

where F and G are the Fourier transforms of the images f and g. Calculating the cross-power spectrum of the two images,

$$C(u,v)=\frac{F(u,v)\cdot G(u,v)^*}{\left|F(u,v)\cdot G(u,v)^*\right|}=\exp(2\pi j(ux_0-vy_0))$$

and then calculating the inverse Fourier transform of the cross-power spectrum produces an image that should have a maximum response at the point corresponding to the translation vector (x_0, y_0). In perfectly matched images, the maximum value should be one. In the presence of noise, the maximum value will not be exactly one, but it should still be a maximum at the correct value of the translation vector. The size of the maximum could also be used as a similarity measure to act as an alternative to mutual information and cross correlation – but this is not considered in this paper.

3 Geometric Correction of Image Motion

For simplicity, the motion of the camera between images is assumed to generate a two-dimensional affine transformation (consisting of a translation, rotation and global scaling). This type of transformation can be represented by a 3×3 matrix in homogeneous image coordinates:

$$T=\begin{pmatrix}1&0&0\\0&1&0\\t_x&t_y&1\end{pmatrix}$$

$$S=\begin{pmatrix}1&0&0\\0&1&0\\-\frac{i_x}{2}&-\frac{i_y}{2}&1\end{pmatrix}\cdot\begin{pmatrix}s&0&0\\0&s&0\\0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0\\0&1&0\\\frac{i_x}{2}&\frac{i_y}{2}&1\end{pmatrix}$$

$$R=\begin{pmatrix}1&0&0\\0&1&0\\-\frac{i_x}{2}&-\frac{i_y}{2}&1\end{pmatrix}\cdot\begin{pmatrix}\cos\theta&\sin\theta&0\\-\sin\theta&\cos\theta&0\\0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0\\0&1&0\\\frac{i_x}{2}&\frac{i_y}{2}&1\end{pmatrix}$$

$$\textit{Transform Matrix}=R\cdot S\cdot T$$

where, by convention, the origin of the coordinate system is in the top left of the image, x-horizontal and y-down. This simplification corresponds to the assumption that objects in the field of view are sufficiently far from the camera that the parallax effects are negligible. Each point in the image (x,y) is transformed according to

$$\begin{pmatrix}x'\\y'\\a\end{pmatrix}=R\cdot S\cdot T\cdot\begin{pmatrix}x\\y\\1\end{pmatrix}$$

where the new location is (x', y') and a is a dummy variable.

Generating the transformation from the control points is relatively simple since the set of simultaneous equations formed by the transformation of the control points from image to image may be inverted to derive the transformation parameters: t_x, t_y, s and θ. The main errors to be studied for this paper were errors in the estimated translations, which are of paramount importance when dealing with image stabilisation as they are sensitive to the errors in the rotation and scaling – large translation errors are normally associated with errors in the other parameters.

Once the transformation parameters are known for successive frames of imagery it is a relatively straightforward process to remove the effect of the transformation and to remove the motion of the camera between frames.

4 Results

A number of image registration techniques were examined for the current study, including a number of different image feature detection and association techniques. Most techniques performed well with good-quality, visible-band imagery, but some were found to be erratic when applied to infrared imagery with higher intrinsic noise levels [3]. As a result, the best performing algorithm for each of the main types was assessed further to determine their sensitivity to translations and occlusions. The algorithms selected were: (i) a feature-based technique, (ii) a region-based technique using the maximisation of the mutual information, (iii) a region-based technique using the maximisation of the cross-correlation, and (iv) the Fourier method described above. The techniques (ii), (iii) and (iv) are described above – variants in the approach to maximising the similarity measure employ different optimisation strategies, but the stochastic gradient method described in reference [6] was found to be the most reliable for the imagery used in this study. Several different feature-based methods were tested the most reliable was found to be a technique using a canny edge detection algorithm and a simple corner detection patch filter applied to each pixel in the image. The corner detection filter scans around the edge of the region surrounding the pixel of interest and detects the transitions between edge and non-edge pixels. It uses the properties of the transitions to determine whether the pixel of interest belongs to an 'L' shaped corner or a 'T' shaped corner. Once the corners were detected, they were associated between the two images using the Hu invariant features.

4.1 Comparison of Errors

The performance of each of the four algorithms selected for further study can be seen in figure 2, where the registration error is plotted as a function of the initial misalignment error. The assessment was performed on a set of simulated infrared images, which was based on the specifications of known infrared imagers used in airborne imaging systems [8,9] – typically, the images corresponded to infrared imagers with 3-5 degree fields of view at altitudes above a thousand metres and were 8-bit monochrome, simulated mid-wave infrared images. They were low contrast images, 512×512 pixels in size and the scene-background was based on freely available satellite imagery with a 1-2 metre resolution.

One of the interesting factors found in the assessment of the errors is the large errors found for the mutual information technique. This is in contrast to the results found in reference [6], where the mutual information was found to be a good similarity metric for image registration. The difference here is that the image registration techniques were also allowed to generate scaling corrections, a factor that was excluded – for good reason – from the study described in reference [6]. With the scaling parameter in the optimisation strategy, the maximisation of the mutual information was found to be unstable. If the scaling is fixed, then the performance of the maximum mutual information is far better, but not significantly better than that found for the maximum cross-correlation technique – so the cross-correlation technique was preferred for the current application.

The performance of the other three techniques was found to be relatively good for small translations of the images (and the Fourier technique proved to be extremely good for a wide range of image translation errors). This is the regime that would be expected for image stabilisation applications. However, after further testing, it was found that the feature-based technique proved to be unreliable when applied to images with structured and unstructured occlusions, of the type anticipated from the exhaust plumes and seen in real airborne infrared imagery.

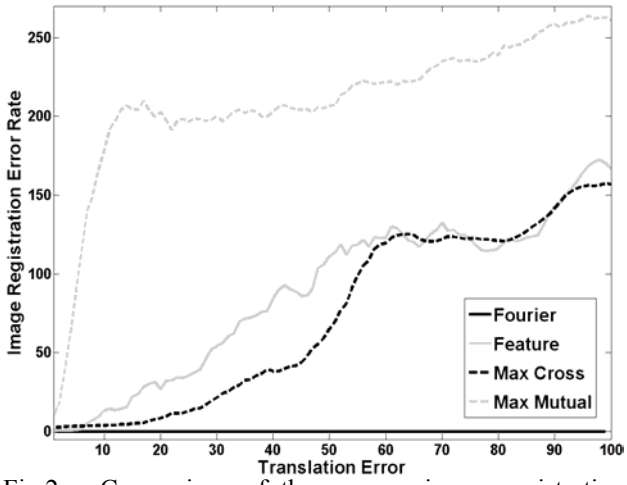

Fig.2 – Comparison of the average image registration errors (in pixels) for four of the techniques studied – including feature-based, region-based and Fourier image registration techniques.

4.2 Region-based methods

The performance of the maximum cross-correlation technique to occlusion by unstructured (randomly distributed) clutter is shown in Figure 3. The performance proved to be extremely robust in the presence of occlusions which cover up to 30-40% of the available image area. Given sufficient structural content in the underlying background scene, the cross-correlation proves to be an extremely good similarity metric even in the presence of occlusion. The need for structural content in the image may appear to be a little restrictive, but a moment of thought shows that such a requirement is a necessary condition for any image registration technique. Fortunately, at the altitudes considered for the current study, the number of ground features visible in most infrared images is relatively large. Even in deserted regions, there are often small tracks and divisions between regions of vegetation that are sufficient to generate a reasonably good correlation.

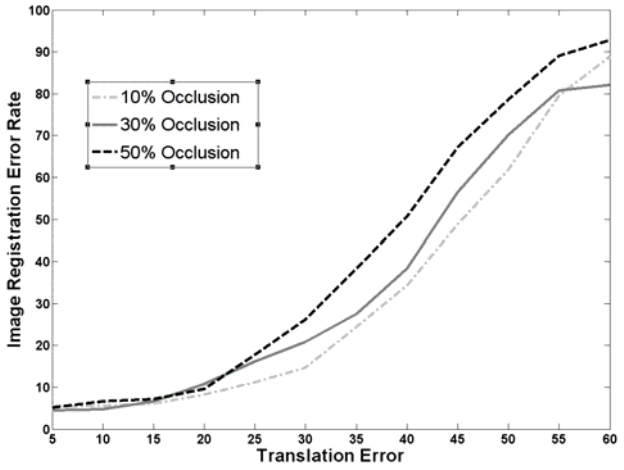

Fig.3 – Comparison of average image registration errors (in pixels) for best performing region-based image registration technique (maximisation of image cross-correlation, using a stochastic gradient optimisation method) with different levels of occlusion.

The robustness of the maximum cross-correlation technique is largely due to the global nature of the correlation process (i.e. it is a region-based similarity measure rather than localised in a few places, like the feature based methods). The main potential problem with using a correlation method is that the structured clutter/exhaust plume is sufficiently slowly varying to give rise to spurious correlations by matching the clutter form image to image rather than matching the background scene. For a relatively fast moving aircraft and a normal frame rate (approx. 50-60 frames per second), this was not observed to be a major problem in the real infrared image data that was available. A source of potential image clutter that could cause such effects would be clouds close to the flight level of the imager. However, a cloud that was too close would pass by the field of view of the imager very quickly and one that was too far away would not move significantly against the background scene from frame to frame and again would not be a problem. This is a problem that requires further study and access to a larger database of infrared images.

4.3 Fourier methods

The Fourier method proved to be the most accurate of the techniques investigated for this paper – see Figure 2. The Fourier method is insensitive to translational errors. It was, after all, designed for exactly this problem and makes use of the fact that a translation in position is equivalent to a phase shift in the corresponding frequency space. By finding and estimating this phase shift is a simple matter to find the corresponding translation. As a result, there is no requirement for the type of sophisticated optimisation search strategy used in the region based methods. The position of the maximum in the inverse Fourier transform of the cross-power spectrum of the two images is sufficient. This is not an entirely fair comparison since the Fourier method cannot generate rotations and scaling parameters in the simple form used here, but generalisations exist to extend it further. However, the simple form used here was sufficient for the present study.

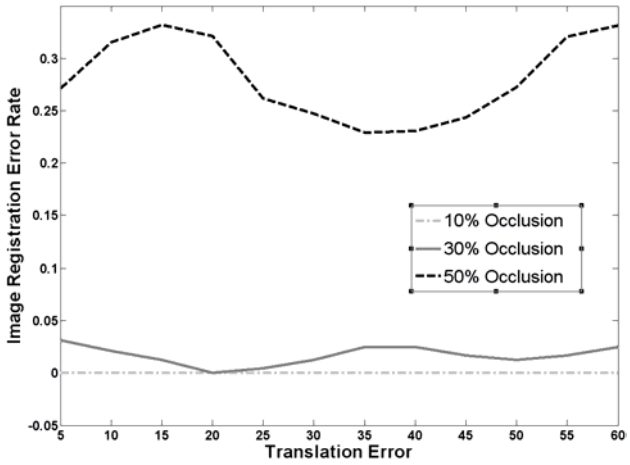

Fig.4 – Comparison of average image registration errors (in pixels) for basic Fourier registration with different levels of occlusion.

Adding in clutter and occlusions into the images the Fourier method proved to be significantly more robust even than the maximum cross-correlation method. The errors for 10%, 30% and 50% occlusions are an order of magnitude less than those

found for the best of the other techniques. Figure 4 shows the average errors as a function of the initial translations. Even with 50% of the image covered by clutter, the match is still very good, and is still approximately independent of the initial translation error. In fact, the image registration errors only become significant when the percentage occlusion rises to about 70-80% for unstructured clutter. This is an extremely robust algorithm and in this extreme regime it is only just possible to distinguish similarities between the two images 'by eye'.

The main potential source of error for this technique is the same as for the maximum cross-correlation – i.e. false correlations due to structured clutter or occlusion. In the case of the maximum cross-correlation the risk is one of spatial correlations causing false registrations. As discussed, these are relatively easy to pick by eye and to predict their likely occurrence based on the spatial relationship between the clutter, the background and the camera. The situation for the Fourier technique is less clear because it is the spatial and the frequency characteristics of the clutter and background that will determine whether any false peaks appear in the cross-power spectra of two images. In particular, if the clutter has one or more dominant frequencies, it is possible that the phase shift and the translations could be erroneous.

5 Conclusions

This paper has considered the use of image registration techniques for the stabilisation of imagery from an airborne infrared camera. Of particular interest is the robustness of the image registration process to the types of noise that can be found in typical infrared imagery: such as low image contrast, exhaust plumes and other clutter that might obscure or occlude part of the background image, and vary from frame to frame.

A number of techniques were investigated to derive the correct two-dimensional affine transformation to correct for motion of the camera between successive frames of imagery. The techniques included standard feature-based, region-based and Fourier methods. It was found that the most reliable for the registration of infrared imagery in the presence of common noise sources and low image contrast were Fourier registration and region-based registration. For region-based registration, a maximum cross-correlation technique was preferred because the maximum mutual information method tested was found to have a high sensitivity to errors in the image scaling parameter. In the presence of significant levels of structured and unstructured clutter, both the maximum cross-correlation and the Fourier registration methods were robust and the Fourier method had the added advantage that it was found to be relatively insensitive to the initial image translation errors. The Fourier method only failed when the clutter was sufficiently dense that the images could not even be matched 'by eye'.

References:

[1] B.Zitova, J.Flusser, 'Image Registration Methods: A Survey', *Image and Vision Computing*, Vol. 21, 2003, pp.977–1000.

[2] H.G.Krapp, R.Hengstenberg, 'Estimation of self-motion by optic flow processing in single visual interneurons' *Nature,* Vol. 384, 1996, pp.463-466.

[3] J.F.Ralph, M.Bernhardt, 'Smart Imaging in the Infrared' *Contemporary Physics*, Vol.43, 2002, pp.259-272.

[4] M-K.Hu, 'Visual pattern recognition by moment invariants', *IRE Trans. on Information Theory*, Vol. IT-8, 1962, pp. 179-187.

[5] J. P. Lewis, 'Fast Template Matching', *Vision Interface,* 1995, pp. 120-123.

[6] A.Cole-Rhodes, K.L.Johnson, J.LeMoigne, I.Zavorin, 'Multiresolution Registration of Remote sensing Imagery by optimization of mutual information using stochastic gradient', *IEEE Trans. on Image Processing,* Vol.12. No 12., 2003, pp.1495-1511.

[7] P.E.Anuta, 'Spatial registration of multi-spectral and multi-temporal digital imagery using Fast Fourier Transform', *IEEE Trans. on Geoscience Electronics,* Vol. 8, 1970, pp.353–368.

[8] J.F.Ralph, K.L.Edwards, S.W.Sims, 'Contextual Targeting', *Proceedings of SPIE conference 'Signal Processing, Sensor Fusion and Target Recognition XIV'*, Orlando, Florida, April 2005, SPIE Vol., Ed. I.Kadar (2005).

[9] J.F.Ralph, M.I.Smith, J.P.Heather, 'Identification of Missile Guidance Laws for Missile Warning Systems Applications', *Proceedings of SPIE conference 'Signal and Data Processing of Small Targets 2006'*, Orlando, Florida, April 2006, SPIE Vol., Ed. O.E.Drummond (2006).