# Image Registration using Multiresolution Frequency Domain Correlation

Stefan Krüger and Andrew Calway
Department of Computer Science
University of Bristol
Bristol BS8 1UB
[stefan|andrew]@cs.bris.ac.uk

**Abstract**

This paper describes a correlation based image registration method which is able to register images related by a single global affine transformation or by a transformation field which is approximately piecewise affine. The method has two key elements: an affine estimator, which derives estimates of the six affine parameters relating two image regions by aligning their Fourier spectra prior to correlating; and a multiresolution search process, which determines the global transformation field in terms of a set of local affine estimates at appropriate spatial resolutions. The method is computationally efficient and performs well for a range of different images and transformations.

## 1 Introduction

Image registration is an important area of Computer Vision and Image Processing. It involves determining the transformation which will map pixels in one image to their corresponding or matching pixels in one or more related images, where the latter are different views and/or different time frames of the same scene or object, for example. The need for registration may be to enable comparison of the registered images, as in applications such as Photogrammetry or Medical Imaging, where mis-alignment between images needs to be corrected, or to make use of the transformation itself in order to determine properties such as motion, depth or shape in Computer Vision applications [1]. The relationship between the images may correspond to a single global transformation, as in the case of different views of a static planar surface for example, or to a spatially varying transformation field, possibly including discontinuities, as in the case of a non-static scene. In either case, the registration problem is a difficult one and often complicated by such things as occlusions, ambiguous matches and the presence of observation noise or distortion.

The task can be eased by using knowledge of the imaging process to constrain the class of transformation field being estimated. The simplest example is to assume that the field is translational, which is sometimes a reasonable assumption in areas such as motion or depth estimation. A more general approach is to assume that the images are related by a six parameter affine transformation, corresponding to dilation, rotation, shear and translation. For example, such a model is appropriate in the case of a planar surface viewed under the assumption of weak perspective projection, which can be generalised to allowing the field to be piecewise affine, corresponding to viewing a smooth 3-D surface

BMVC 1998 doi:10.5244/C.12.32

under weak perspective. The affine approach gives considerable flexibility in estimating a wide range of transformation fields and has been adopted in a large number of registration techniques (see [1] and [2] for an overview). Within these, the use of Fourier techniques has received considerable attention, with the symmetry and invariant properties of the Fourier domain being exploited to derive affine estimates between images and image regions. Examples include methods based on moments [3], generalised correlation [4]-[6], and other schemes [7][8].

The method described here is also a frequency domain technique. It assumes a piecewise affine transformation field and uses a generalised correlation scheme to estimate the local transformations. The underlying principle is the same as that in [4]-[6]: transform to the frequency domain and align the spectra prior to correlating. The required spectral alignment corresponds to the linear component of the affine transformation and can be estimated by relating the shift-invariant magnitude spectra. An estimate of the translation component then follows from the correlation field. The technique has the advantage of decoupling the linear component search space from that of the translation and has the robustness of a correlation estimator. However, most previous implementations have limited the spectral alignment to a subspace of transformations, such as dilation and rotation [1][4][6], hence restricting its use for image registration. The technique described here is more general in that it is also able to deal with shear alignment and hence yields full six parameter affine estimates (an alternative method for doing this was also proposed in [5]). The alignment process is based on centroids, making it potentially more robust than the approach taken in [5], and can be implemented efficiently, making it preferable to an exhaustive search approach [7]. The principle of the method was originally described in [9] in the context of texture analysis (see also [10]) and was reported in [11] and [12] in the context of affine representations and motion estimation, respectively.

To deal with piecewise affine fields and large translations, the method is incorporated within a multiresolution search process, in which estimates obtained within large regions are used to determine which smaller regions at the ‘next level’ are correlated. This avoids exhaustive searching and ensures that the correlation is performed at a resolution appropriate to the data. The multiresolution version of the affine estimator was originally developed for motion estimation as described in [12].

The purpose of this paper is to report several improvements to the affine estimator and to illustrate its use in the wider context of image registration. Results of experiments on a number of different types of image data demonstrate that the method performs well even in the presence of fairly severe transformations and additive noise.

## 2 Adaptive frequency domain correlation

The basic principle underlying the affine estimator is based on the affine theorem of the Fourier transform [13]. Given two images $x_1(\vec{\xi})$ and $x_2(\vec{\xi})$ related by an affine transformation, ie

$$x_2(\vec{\xi}) = x_1(A^{-1}(\vec{\xi} - \vec{b})) \tag{1}$$

where $A$ is a $2 \times 2$ invertible matrix defining the linear component of the transformation and $\vec{b}$ is the translation vector, then the Fourier transforms (FT) of the two images are related as follows:

$$\hat{x}_2(\vec{\omega}) = |\det A|\, e^{-\jmath\vec{\omega}.\vec{b}}\, \hat{x}_1(A^T\vec{\omega}) \tag{2}$$

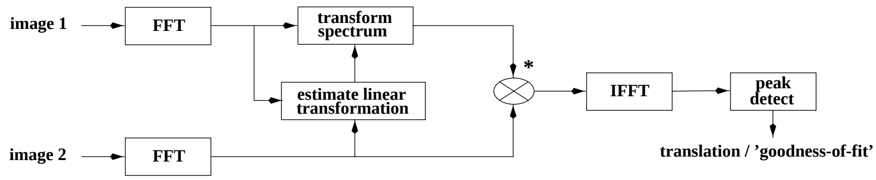



Figure 1: Affine estimation via the frequency domain

where $A^T$ denotes the transpose of A, $\det A$ its determinant and '.' the scalar product. The significance of the above equation is the separation of the six affine parameters in Fourier space: the shift-invariant magnitude spectra are related by the linear component $A$, whilst the translation vector is 'carried' in the phase. This suggests the following three stage algorithm to determine the transformation relating the two images (see Fig. 1):

1. Find the linear transformation $A$ relating the magnitude spectra $|\hat{x}_1(\vec{\omega})|$ and $|\hat{x}_2(\vec{\omega})|$.

2. Align the two spectra by applying the coordinate transformation defined by $A$ to one of the spectra, ie $\hat{x}_1(\vec{\omega}) \to \hat{x}_1(A^T\vec{\omega})$.

3. Correlate by forming the conjugate product of the aligned spectra, ie $\hat{x}_{12}(\vec{\omega}) = \hat{x}_1^*(A^T\vec{\omega})\hat{x}_2(\vec{\omega})/(E_1E_2)$, where $E_i^2 = \int |\hat{x}_i(\vec{\omega})|^2 d\vec{\omega}$, and performing an inverse Fourier transform.

If equation (1) holds, then the resulting correlation field will then be the normalised auto-correlation of $x_2(\vec{\xi})$ translated by the vector $\vec{b}$. If this field is denoted by $\rho(\vec{\xi})$, then it has its peak value at $\vec{\xi} = \vec{b}$, enabling the translation to be determined. Also, the peak value then provides a 'goodness-of-fit' measure, indicating how well the two images match following the estimated transformation.

The advantage of the above algorithm is that the search space for the linear component is decoupled from that for the translation, giving a search over a four parameter space followed by one over two parameters, rather than one over all six affine parameters. Moreover, estimation of the translation via correlation provides robustness in the presence of noise. Of course, the key element is the estimation of the linear component $A$ in order to align the spectra. A robust method for doing this based on centroids is described below.

# 3 Spectral alignment

## 3.1 Centroid pairs

The spectral alignment is achieved by matching centroids computed within angular segments of the magnitude spectra. The use of an angular segmentation follows from the symmetry of the FT and the fact that it is preserved by the invertible transformation, ie the segments are equivariant under the transformation. The centroids are also equivariant and hence provide a means of estimating the transformation: if $\Lambda(\theta_{i1}, \theta_{i2})$, $i = 1, 2$, denote corresponding angular segments between the angles $\theta_{i1}$ and $\theta_{i2}$ in the two spectra as illustrated in Fig. 2, ie the line through the origin at angle $\theta_{2j}$ maps to the line at angle $\theta_{1j}$

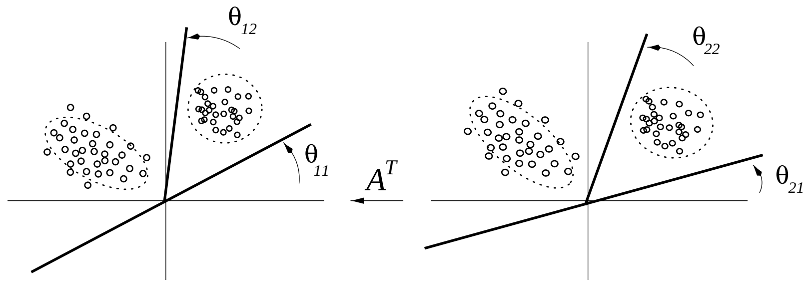



Figure 2: Angular segmentation of the magnitude spectra.

under the action of $A^T$, then the centroids in the segments

$$\vec{\mu}_i(\theta_{i1}, \theta_{i2}) \;=\; \int_{\Lambda(\theta_{i1},\theta_{i2})} \vec{\omega}\,|\hat{x}_i(\vec{\omega})|\,d\vec{\omega} \;\Big/ \int_{\Lambda(\theta_{i1},\theta_{i2})} |\hat{x}_i(\vec{\omega})| d\vec{\omega} \tag{3}$$

are related by

$$\vec{\mu}_1(\theta_{11}, \theta_{12}) \;=\; A^T \vec{\mu}_2(\theta_{21}, \theta_{22}) \tag{4}$$

To determine all four elements of $A$ requires at least two such correspondences. The simplest way of doing this and one which makes full use of the available data is to use two segments which partition an oriented half plane, ie $\Lambda(\theta_{i1}, \theta_{i2})$ and $\Lambda(\theta_{i2}, \theta_{i1} + \pi)$, as illustrated in Fig. 2. This gives a 'centroid pair' for each spectra and hence two vector equations such as that in eqn (4), enabling all the elements of $A$ to be found.

## 3.2 Equivariant segments

Given two spectra to be aligned, it remains to determine the segment angles $\theta_{i1}$ and $\theta_{i2}$ in order to compute the centroids. Two criteria can be identified:

1. That each segment should contain significant energy corresponding to distinct image structure in the spatial domain; and

2. That the segments be defined in terms of a measure which is invariant under the linear transformation.

The first of these ensures that the centroids 'latch on' to significant image structure (which 'carries' the registration information), while the second ensures that corresponding segments are found in the two spectra, ie that the segments are equivariant under the linear transformation. In [9]-[12], the segments were identified as those for which the sum of the segment variances was a minimum over all possible half-planes and segments. This satisfies the first criteria, although fails to satisfy the second for all possible transformations (in particular those involving shear). A related approach but one which satisfies both criteria is to seek the segments in each spectra such that the sum of the determinants of their second moment matrices is a minimum, ie defining the second moment matrix of the segment $\Lambda(\theta_{i1}, \theta_{i2})$ as

$$M_i(\theta_{i1}, \theta_{i2}) \;=\; \int_{\Lambda(\theta_{i1},\theta_{i2})} \vec{\omega}\,\vec{\omega}^T\,|\hat{x}_i(\vec{\omega})|\,d\vec{\omega} \tag{5}$$

then the angles $\theta_{i1}$ and $\theta_{i2}$ are sought which minimise

$$m(\theta_{i1}, \theta_{i2}) \;=\; \det M_i(\theta_{i1}, \theta_{i2}) \;+\; \det M_i(\theta_{i2}, \theta_{i1} + \pi) \tag{6}$$

It is readily shown that $m(\theta_{i1}, \theta_{i2})$ is invariant to within a scale factor under linear transformation [14], and hence the corresponding segments identified in the two spectra will be related by Equation (4) (assuming the spectra are similarly related). Moreover, since $\det\ M_i(\theta_{i1}, \theta_{i2})$ represents the degree of variation about the centroid in the respective segment (it corresponds to the product of the variances along the principle axes as represented by the ellipses in Fig. 2), minimising eqn (6) results in segments which divide the half-plane spectrum into distinct energy clusters, as required by the first criterion above.

### 3.3 Implementation and local searching

The above affine estimator can be implemented efficiently in practice. Using discrete spectra for the images or image regions (see Section 4), partial summations formed over discrete angular segments enable fast computation of the centroids and second moment matrices and hence minimisation of eqn (6). However, the symmetry of the FT means that the correspondence between the centroids in the two spectra is ambiguous and can only be resolved by correlating. In theory there are eight possible pairings (and hence eight transformations to test), although this can be reduced to four by requiring $\det A > 0$, which makes the reasonable restriction that the possible transformations do not involve a reflection. The transformations can be tested by first aligning the spectra (using bilinear interpolation) and then correlating to determine which gives the best fit (Fig. 1).

The performance of the algorithm can be significantly improved, particularly in the presence of severe transformation or noise, by employing a limited search to better align the spectra following initial estimation of the matrix $A$. This improves both the estimate of $A$ and of the translation $\vec{b}$, since the latter is critically dependent on correct alignment of the spectra. The approach adopted is a two stage process in which a search is performed over a set of transformations corresponding to vectors in the vicinity of one pair of corresponding centroids while keeping the other pair fixed. A given transformation is tested by computing the inner product between one magnitude spectrum and a transformed version of the other within a small window about the respective centroids. This is reasonable since it is assumed that the search is about an initial estimate which is close to the required transformation (and hence the spectra are already closely aligned) and avoids computing a full correlation. The transformation giving the largest inner product is then selected. A final correlation can then be computed and a sub-pixel estimate of the translation obtained using standard cubic interpolation.

## 4 Multiresolution decomposition and searching

The above estimation algorithm is effective when applied to image regions of an appropriate size: large regions are likely to contain too many features, preventing robust angular segmentation of the spectra; whilst small regions may contain too few or no features, preventing registration. Moreover, in the case of a piecewise affine transformation field, the estimator needs to be applied within regions which match the affine model. However, adopting a region based approach means that the correspondence problem needs to

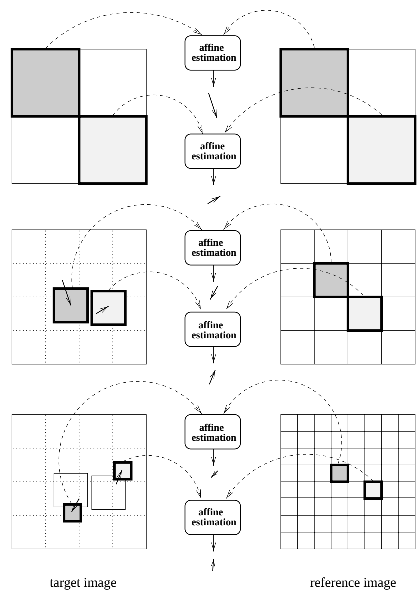
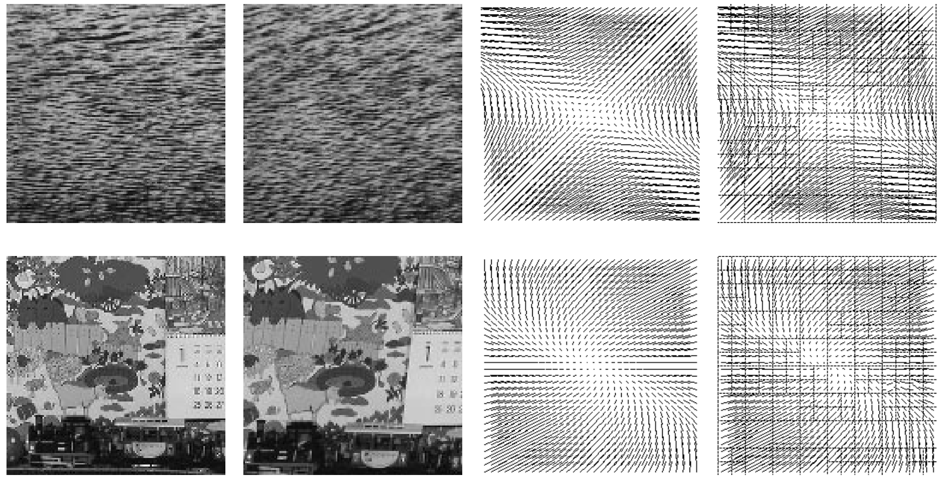



Figure 3: Multiresolution search procedure defined with respect to a quadtree tessellation of the reference image.

be solved - determining which regions in the two images need to be registered. Both of these criteria can be addressed by employing a multiresolution search procedure in which estimates obtained within large regions are used to predict the correspondence of smaller regions at the ‘next level’, with regions being ‘selected’ if the goodness-of-fit measure is sufficiently high. This yields a multiresolution decomposition of the transformation field defined in terms of a set of local affine transformations.

The search procedure operates with respect to a quadtree tessellation of one of the images (the reference image) as illustrated in Fig. 3. Starting at some large region size (typically $64 \times 64$ pixels in the experiments), corresponding regions are registered using the affine estimator. These estimates then determine which regions each of four ‘child regions’ below are registered with in the other image, ie if $\vec{\xi}$ is the position vector of the centre of a child region defined with respect its parent region centre, then the region is registered with the region centred at $A\vec{\xi} + \vec{b}$, where $A$, $\vec{b}$ is the affine estimate obtained at the parent region. In some applications, it is also beneficial to use a four parent quadtree, in which case a given child region can be registered with up to four possible corresponding regions in the other image. This is often necessary in the vicinity of discontinuities in the transformation field, eg at motion boundaries.

The search procedure continues down each ‘branch’ of the quadtree until the goodness-of-fit measure, ie the correlation peak, at a given region is sufficiently high and the affine estimate is consistent with those of its child regions. The latter is tested by transforming each child spectrum according to the parent estimate and forming the inner product with the spectrum of its corresponding region in the other image [12]. If the parent correlation peak and all four child inner products are above some threshold then the region is ‘selected’ and the search procedure terminates for that branch.



(a) (b) (c) (d)

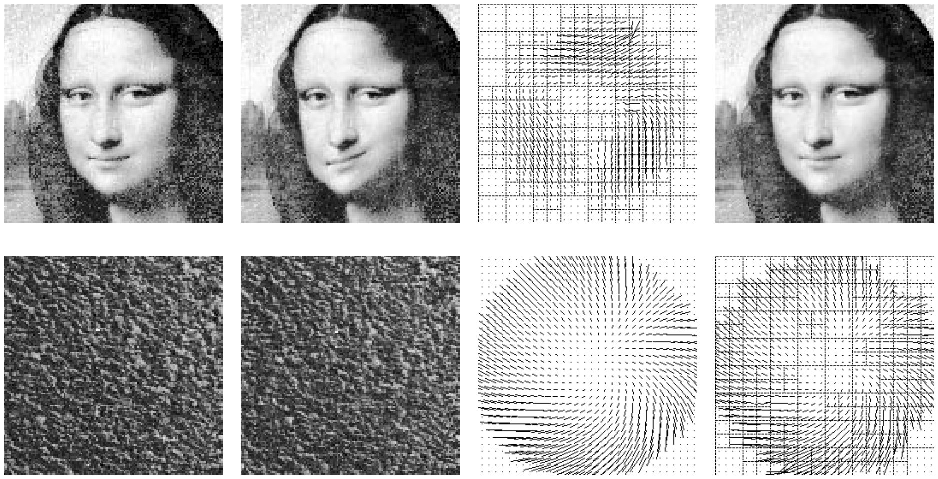
Figure 4: Single global transformations: (a) original images; (b) warped images (see text); (c) true transformation field; (d) estimated transformation field.

# 5 Experiments

To illustrate the effectiveness of the algorithm, results are presented of registration experiments involving different image types and transformation fields. The local spectra for the regions were computed using the multiresolution Fourier transform [15], employing a 2-D separable Hamming window with 50% overlap between the regions. The images were pre-filtered using a high-pass filter with bandwidth proportional to the region size in order to minimise any bias in the correlations (this can be implemented efficiently using a standard pyramid scheme). In addition, the spectral alignment was implemented using the squared magnitude spectra which was found to give more reliable angular segmentation.

The first two experiments involve a single global transformation such as that resulting from a planar surface moving in 3-D. The images in Fig. 4a were warped using the matrices $[[1.0, 0.2], [0.1, 0.9]]$ and $[[0.9, 0.1], [0.0, 0.9]]$ (origin at the centre of the image) to give the images in Fig. 4b. These correspond to a shear along both axes followed by a dilation along the vertical axis and an isotropic dilation followed by a shear along the horizontal axis, respectively. The true and estimated transformation fields are shown in Figs. 4c-d. The grid on the estimated field indicates the selected regions in the multiresolution decomposition and each region has an affine estimate as shown by the arrows. The estimated transformation fields at pixel resolution (obtained using bilinear interpolation) are very close to the known fields, with over 80% of the error magnitudes being $< 1$ pixel.

The second two experiments involve a spatially varying transformation field which is approximately piecewise affine. Figs. 5a-b show the original and warped images. Fig. 5c (top) shows the estimated warp field for the Mona Lisa image and Fig. 5d (top) is the reconstruction of the original from the warped image using the estimated warp field. The reconstruction is very close to the original with only a small number of errors. Fig. 5c (bottom) shows the true warp field applied to the Cork texture image and Fig. 5d (bottom) is the estimated field. Again the estimates are very close to the known warp field. In fact,



(a) (b) (c) (d)

Figure 5: Spatially varying warp fields: (a) original images; (b) warped images; (c) top - estimated warp field (Mona), bottom - true warp field (Cork); (d) top - reconstructed image, bottom - estimated warp field (Cork).

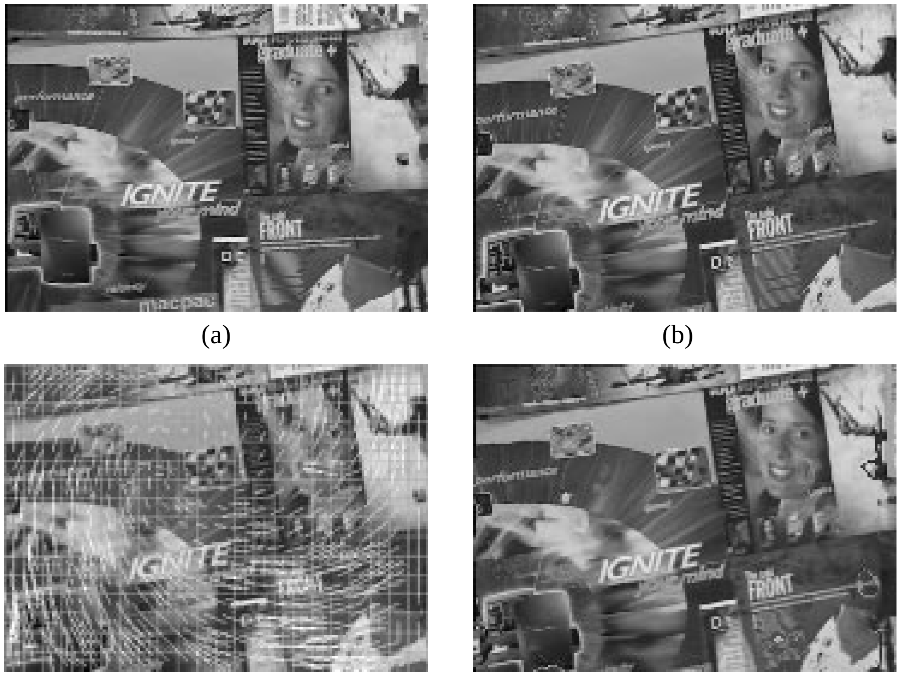
for this example, the images were corrupted by additive, uncorrelated Gaussian noise to give a SNR of 10dB. This illustrates the robustness of the scheme in the presence of noise.

The final experiment involves two frames from a video sequence of a notice board in which the camera was moved towards the board while being rotated. The two frames are shown in Fig. 6a-b and the estimated transformation field (corresponding to the camera motion) is shown in Fig. 6c. The movement of the camera is clearly apparent from the field and its validity is confirmed by the reconstruction of the frame in Fig. 6b from the frame in Fig. 6a as shown in Fig. 6d.

# 6 Conclusions

A general-purpose technique for image registration has been presented. It has several key elements. The spectral alignment using centroid pairs facilitates the use of correlation with its associated robustness for estimating full six parameter affine transformations between local regions. Its incorporation into a multiresolution search procedure enables the technique to adapt to the data, ensuring that the estimator is applied to regions with sufficient and distinct structure to enable registration. It also means that the technique can deal with piecewise affine transformation fields. Moreover, the resulting multiresolution decomposition provides a concise representation of the field which will have advantages for any post-processing stage, eg in the identification and tracking of objects or regions in a motion analysis application [12]. The technique has been shown to perform well for a range of different images and transformations, even in the presence of additive noise.

Nevertheless, the approach is not without its shortcomings. In particular, when faced with regions containing a discontinuity in the transformation field, eg in the vicinity of a motion boundary, the performance of the algorithm is sometimes unpredictable, which is




(a) (b)

(c) (d)

Figure 6: Camera motion example: (a)-(b) original frames 1 and 2; (c) estimated transformation field; (d) reconstruction of frame 2 from frame 1.

not surprising given that the underlying model is that of a single affine transformation. A related problem is the influence of aperture effects, in which single oriented features have an ambiguous transformation field. This is dealt with to some extent by the multiresolution search which will tend to force the estimation to be performed within larger regions containing more than one feature. However, this cannot always be guaranteed. Work is in progress to address both of these issues.

**Acknowledgements:** Thanks to Roland Wilson and Tao-I Hsu. The work was supported by the UK EPSRC.

# References

[1] L.G.Brown, "A survey of image registration techniques", *ACM Computing Surveys*, vol. 24(4), pp. 325-376, 1992.

[2] J.Sato and R.Cipolla, "Image registration using multi-scale texture moments", *Image and Vision Computing*, vol. 13(5), pp. 341-353, 1995.

[3] B.J.Super and A.C.Bovik, "Shape from texture using local spectral moments", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17(4), pp. 333-343, 1995.

[4] E.De Castro and C.Morandi, "Registration of translated and rotated images using finite Fourier transforms", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9(5), pp. 700-703, 1987.

[5] L.Lucchese, G.M.Cortelazzo, and C.Monti, “Estimation of affine transformations between image pairs via Fourier transform”, *Proc. IEEE International Conference on Image Processing*, Lausanne, pp. 715-718, 1996.

[6] B.S.Reddy and B.N.Chatterji, “An FFT-based technique for translation, rotation, and scale-invariant image registration”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5(8), pp. 1266-1271, 1996.

[7] J.Krumm and S.Shafer, “Shape from periodic texture using the spectrogram”, *Proc. IEEE Conf on Computer Vision and Pattern Recognition*, Urbana Champaign, pp. 284-289, 1992.

[8] J.Malik and R.Rosenholtz, “Computing local surface orientation and shape from texture for curved surfaces”, *International Journal of Computer Vision*, Vol.23(2), pp. 149-168, 1997.

[9] Tao-I.Hsu, A.D.Calway and R.Wilson, “Texture analysis using the multiresolution Fourier transform”, Proc. Scan. Conf. on Image Analysis, Tromso, pp. 823-830, 1993.

[10] Tao-I.Hsu and R.Wilson, “A two-component model of texture for analysis and synthesis”, to appear in *IEEE Trans. Image Processing*, 1998.

[11] A.D.Calway, “Image representation based on the affine symmetry group”, *Proc. IEEE International Conference on Image Processing*, Lausanne, pp. 189-192, 1996.

[12] S.A.Krüger and A.D.Calway, “A multiresolution frequency domain method for estimating affine motion parameters”, *Proc. IEEE International Conference on Image Processing*, Lausanne, pp. 113-116, 1996.

[13] R.N.Bracewell, K.-Y.Chang, A.K.Jha, and Y.-H.Wang, “Affine theorem for two-dimensional Fourier transform”, *Electronics Letters*, vol. 29(3), p.304, 1993.

[14] J.Flusser and T.Suk, “Pattern recognition by affine moment invariants”, *Pattern recognition*, vol. 26(1), pp. 167-174, 1993.

[15] R.Wilson, A.D.Calway, and E.R.S.Pearson, “A generalised wavelet transform for Fourier analysis: the multiresolution Fourier transform and its application to image and audio signal analysis”, IEEE Trans. on Information Theory, 38(2), pp. 674-690, 1992.