Image Registration Methods for Resolution Improvement

D.G. Bailey and T.H. Lill

Institute of Information Sciences and Technology, Massey University,
Private Bag 11222, Palmerston North, New Zealand
D.G.Bailey@massey.ac.nz

Abstract

Image registration is a key step in combining multiple independent low-resolution images to give a single high-resolution image. Four sub-pixel registration methods are compared in terms of sub-pixel accuracy in the presence of low to moderate levels of noise. While traditional correlation and differencing methods are relatively insensitive to noise, their accuracy is limited to 5 – 20% of a pixel, particularly in the presence of significant fine detail. Using the phase gives significantly better results (1 – 2% of a pixel) but is more sensitive to noise. A new registration method based on predictive interpolation is described. This gives similar performance to the using the phase, but requires considerably less computation. With low noise and significant fine detail, either the phase predictive interpolation are suitable, with either giving more than adequate performance in this application.

Keywords: *resolution, reconstruction, registration, correlation, phase, interpolation*

1 Introduction

Image fusion involves the combining of data from a number of image sources in such a manner that the output image contains more information than that contained in any of the individual images. The image sources are often different in their properties or characteristics. An important use of image fusion is to trade temporal resolution for spatial resolution, by combining several low-resolution images to construct a high-resolution image [1-3].

The pixels of an image provide a set of samples of the world. The sampling density, or the spacing between the pixels, limits the achievable resolution. However if a series of images is captured, each with the samples in slightly different locations, the combined sample density is higher that that of any single image. Constructing a high-resolution image consists of three stages. First, all of the individual low-resolution images must be registered to one another (correcting if necessary for any image dependent distortion [4]). Next, the ensemble is resampled using a single high-resolution sampling grid. Finally, an inverse filter is applied to compensate for limitations in the image capture process [4].

For the ensemble to contain more information than any one of its constituent images, a single image must not contain all of the information. If each low-resolution image was sampled at the Nyquist rate (or higher) then each image contains sufficient information to be able to recover the original data exactly by using sinc interpolation. This means that any single image would be able to provide all of the information required for reconstruction at any desired resolution. Having multiple independent images would not provide an improvement. Therefore, for multiple images to provide additional information, the sample frequency for each individual image must be below the Nyquist rate, and the images be subject to aliasing. The process of constructing a higher resolution image untangles the aliased information, so that the output image contains more information than that available from any of the individual input images.

An important step within the reconstruction process is to register the low-resolution images with one another so that they may be combined. It is this registration step that this paper addresses.

2 Registration methods

There is a range of registration methods that may be used. The traditional approach to sub-pixel registration is to interpolate each of the low-resolution images to the desired resolution, and then perform a registration at the higher resolution to the nearest pixel. There are two problems with this approach (apart from the increased computational complexity of operating at a higher resolution). First, the chosen higher resolution limits the accuracy of the registration (unless sub-pixel resolution techniques are also applied at the higher resolution). Second, since the images contain aliased information, the interpolation filter effectively leaves this information tangled. Subsequent resampling of the interpolated images will not recover this aliased information. Therefore it is preferable to perform the registration between the low-resolution images.

The key requirement is that the registration method must be capable of sub-pixel accuracy, maintain the accuracy in the presence of aliasing, and a low level of noise, and not rely on the presence of particular objects or other structure within the image. The three main classes of image registration algorithm that fit these requirements are correlation methods, difference methods, and phase methods. A new predictive interpolation method for registration is also presented and compared.

2.1 Correlation methods

Correlation is one of the most popular methods for registering images [5]. There are several variations of correlation, but the basic method is to multiply one image with an offset second image on a pixel by pixel basis. The product is then accumulated and normalised by the overlap area and average pixel value to give the correlation for the particular offset. Let the images be $f(x,y)$ and $g(x,y)$, then the unnormalised correlation $c(i,j)$ is given by

$$c(i,j) = \sum_x \sum_y f(x,y) g(x-i, y-j) \tag{1}$$

This is repeated for a range of different offsets, (i,j), with the offset giving the maximum correlation corresponding to the best match. If the offset is known to be small, the correlation may be performed directly using equation (1), however if the offset is unknown or large, it may be more efficient to perform the correlation in the frequency domain. If $F(u,v)$, $G(u,v)$ and $C(u,v)$ are the Fourier transforms of $f(x,y)$, $g(x,y)$ and $c(i,j)$ respectively, then

$$C(u,v) = F(u,v) G^*(u,v) \tag{2}$$

where * denotes the complex conjugate. The maximum value of $c(i,j)$ gives the offset to the nearest pixel - to register with sub-pixel accuracy, an interpolation surface is fitted to the correlation values, and the interpolated maximum found. Around the maximum, the expected shape of the peak is a pyramid, with a width of twice the size of the features in the image. For large objects without a lot of fine detail, the pyramid can easily be detected and gives an accurate registration. However, since this application is to improve the resolution of the images, significant fine detail is to be expected, and the registration would be less accurate. Therefore only local information should be used. The interpolation process is illustrated for the 1-dimensional case in figure 1. Using the definitions of figure 1, the position of the correlation peak is therefore:

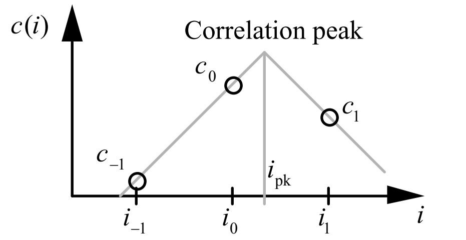


Fig 1: Locating the correlation peak in 1D.

$$i_{\mathrm{pk}} = i_0 + \frac{c_1 - c_{-1}}{2(c_0 - \min(c_1, c_{-1}))} \tag{3}$$

This interpolation is performed in both directions giving the position of the correlation peak to sub-pixel accuracy in two dimensions. However, if the image contains significant fine detail (smaller than a pixel), the size of the pyramid is also small and the accuracy of the interpolation deteriorates, reducing the accuracy of the consequent sub-pixel registration.

2.2 Difference methods

Difference methods are based on the absolute difference between a reference image and an offset second image on a pixel by pixel basis. The difference is accumulated and normalised by the overlap area. Like the correlation methods, the difference is found for a range of offsets, only this time the best match corresponds to the offset with the minimum difference. The unnormalised difference $d(i,j)$ is given by

$$d(i,j) = \sum_x \sum_y \left| f(x,y) - g(x-i, y-j) \right| \tag{4}$$

The expected shape of the local minimum is also pyramidal so an interpolation function similar to equation (3) is used to give the registration to sub-pixel accuracy. The width of the pyramid is twice the image feature size, so difference methods suffer from similar problems to correlation methods in determining the sub-pixel offset.

2.3 Phase based methods

Phase methods of image registration work on the principle that an offset in the image domain gives a linear phase response in the frequency domain [6]. The reference image and second image are Fourier transformed and the phases subtracted. A planar surface is then fitted to the phase difference image and the offset calculated from the phase tilt [3]. If $F_\phi(u,v)$ and $G_\phi(u,v)$ are the frequency domain phase images of $f(x,y)$ and $g(x,y)$ respectively, then the expected phase difference is

$$F_\phi(u,v) - G_\phi(u,v) = x_0 u + y_0 v \tag{5}$$

where (x_0, y_0) is the offset between the two images. In practise, the images need to be windowed before Fourier transformation to prevent any discontinuity at the edge of the image from dominating the phase response (since the FFT assumes that the image is periodic). After the phase difference image is calculated, it needs to be "unwrapped" by the addition of $\pm 2n\pi$ at each point to make the phase as planar as possible. A least squares fit of a plane to the unwrapped phase difference image gives the offset. A better fit is obtained if the individual phase points are weighted by the corresponding magnitude because the phase of those frequencies with low magnitude are more affected by noise. Finally, only the lower frequency portion of the phase image is used because the higher frequencies are more affected by aliasing.

2.4 Predictive interpolation method

Another method that may be used for registration is predictive interpolation. This method attempts to predict the pixel values of the second image as a function of the pixel values in the reference image. The parameters of the linear prediction equation

$$h(x,y) = g(x-i, y-j) = A_{00} f(x,y) + A_{01} f(x,y+1) + A_{10} f(x+1,y) + A_{11} f(x+1,y+1) \tag{6}$$

are determined using a least squares fit, subject to the constraint that

$$A_{00} + A_{01} + A_{10} + A_{11} = 1 \tag{7}$$

The offset to sub-pixel accuracy is then given by

$$(x_0, y_0) = (i + A_{10} + A_{11}, j + A_{01} + A_{11}) \tag{8}$$

The prediction function is effectively interpolating a test image, $h(x,y)$, from the reference. The parameters that give the best match (in a least squares sense) between the test image, $h(x,y)$, and the second image, $g(x,y)$, give the offset between the images. In determining the prediction coefficients, a weighted least squares is used, with the individual pixels weighted by the local variance. This prevents featureless regions within the image from giving meaningless values.

Using equation (6) directly requires that the two images already be registered to the nearest pixel ($i = j = 0$). However, if the calculation is repeated for a range of offsets, the offset providing the minimum prediction error will give the best fit to the pixel level.

2.5 Other registration methods

The registration methods described so far make no explicit assumptions about the nature of the images being registered apart from the fact that there is sufficient detail within the image to give a meaningful registration. In circumstances where the image is known in advance to be of a particular form, other high-level registration methods may be used.

Where the image can easily and reliably be segmented into a number of separate regions (for example if the image contains one or more isolated objects), the centre of gravity of the segmented regions may be determined to sub-pixel accuracy. In general, the size of the regions and the accuracy of the segmentation limit the accuracy of the registration.

If the image contains a number of straight edges, these may be detected and parameterised to sub-pixel accuracy. Again, the length of the edge and the accuracy of detection limit the accuracy of registration.

These approaches have not been included in this comparison because they rely on some predetermined form of structure within the images. If the image consists predominantly of features smaller than a pixel, structure based methods cannot be used to determine the location of these features.

3 Evaluating the registration

To compare the registration methods, it is necessary to work with a set of images where the offsets are known in advance. This was accomplished by taking a high-resolution image, and subsampling it to produce a series of low-resolution images with predetermined offsets. The registration methods were then used to measure the offsets, and these results were compared with the known offsets to determine the registration error. In the tests here, the original images were 512x512 pixels. These were subsampled by a factor of three by first smoothing the images using a 3x3 average filter and then selecting every 3rd pixel in the horizontal and vertical directions. This simulates the capture of an image using a lower resolution camera with sensing elements 3 times larger than the original image. This process gave a set of nine 171x171 pixel low-resolution images that were offset from one another by multiples of 1/3 pixel. The 9 low-resolution images allowed 36 individual relative offsets to be determined. The RMS error between these 36 measurements and the known offsets gave an estimate for the registration accuracy for that particular image.

The sample images that were used are shown in figure 2. The images were chosen for their varying level of detail, from a fully structured image (A), two general images (B and C) with varying levels of detail, a high resolution text image (D) and a low resolution text image (E) with inadequate resolution to be able to read the text. The final image is expected to present a considerable challenge because the fine detail within the image will result in considerable aliasing. Since the primary purpose of this study is to improve the image resolution, the success on image E is desirable since it indicates the extent to which a high-resolution reconstruction is possible.

To investigate the sensitivity of the different registration methods to noise, random Gaussian noise was independently added to each of the low-resolution images. The RMS error was measured 10 times at each noise level and the results averaged to reduce the effects of a specific noise pattern.

In performing the registration, the central 128x128 block from each of the low-resolution images was used. This size was chosen for its convenience for the FFT used by the phase registration.

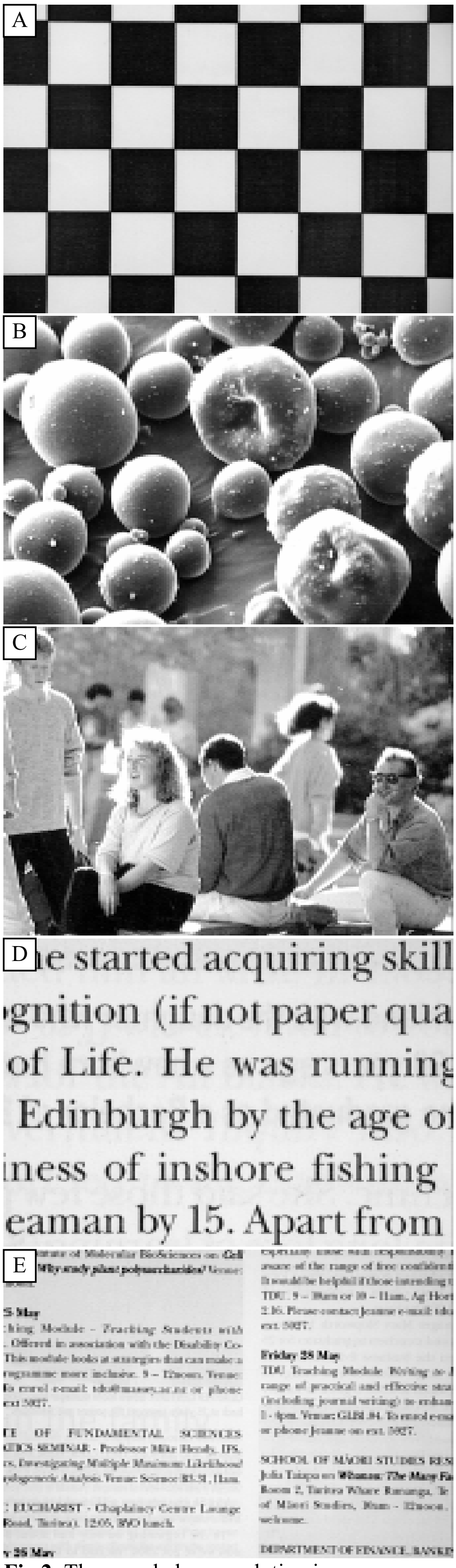


Fig 2: The sample low-resolution images.

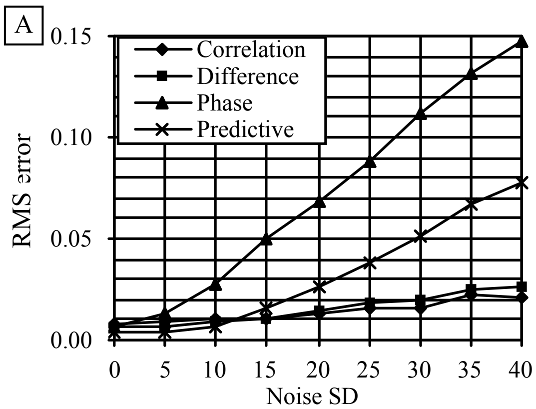
The pair-wise registration between the 9 low-resolution images gives 36 separate registration measurements. Of these, there are only 8 independent offsets. Therefore, if the registration errors are random, the average error may be reduced by enforcing consistency between the separate measurements. Let the low-resolution images be f_i and the measurement of the offset between f_i and f_j be M_{ij} where $i < j$. The first image f_1 can arbitrarily be chosen as the reference. It can then be shown that the average offset of each of the other images from the reference $\overline{M}_{1i}$ is given by

$$\overline{M}_{1i} = \frac{1}{N}\sum_{j<i} M_{ji} - \sum_{i<j} M_{ij} + \sum M_{1j} \tag{9}$$

4 Results and discussion

Table 1 shows the effect of enforcing consistency between the pair-wise registrations. It tended to have less of an effect on the phase registration results, which were more consistent. The most notable improvement was in the results for the predictive interpolation method. In the presence of significant noise, the improvement tended toward the expected factor of $\sqrt{8/36}$ (0.48).

Figure 3 compares the performance of the four registration algorithms on each of the 5 test images. On image A, the phase registration method did not perform well in the presence of noise. This is because



Method	No noise	Noise sd > 15
Correlation	0.5-0.8	0.5-0.8
Difference	0.5-0.8	0.5-0.8
Phase	0.7-1.0	0.5-0.9
Predictive	0.1-0.3	0.4-0.5

Table 1: Reduction in RMS error from enforcing consistency between individual offsets.

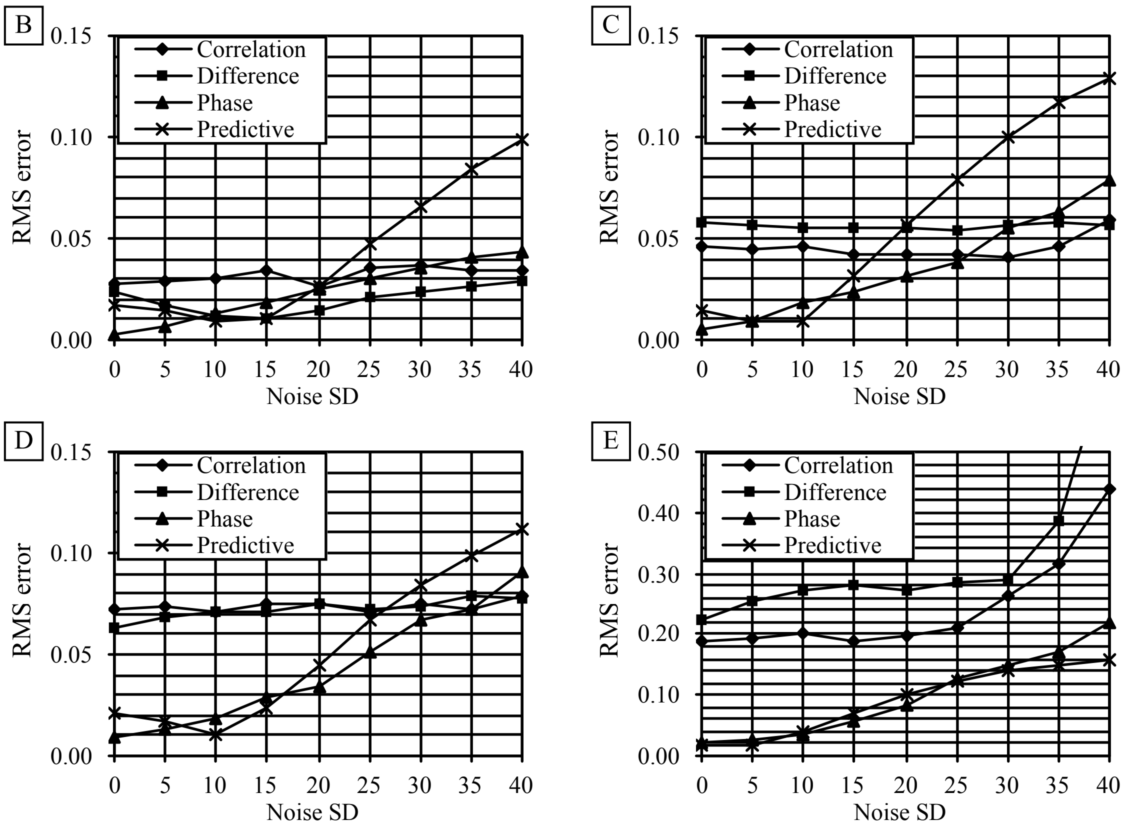


Fig 3: Registration error as a function of noise level of the 4 algorithms applied to each of the test images.

only a few frequencies contain significant information and when noise was added, it perturbed the fitted phase. The predictive method was also affected by the noise because there are few edges within the image, and the prediction is weighted by the variance. Images B, C and D show that the phase method performs well with low to moderate levels of noise, but deteriorates linearly with significant noise. An interesting feature of the predictive method is that the accuracy improves with the addition of a small amount of noise. The noise must have a dithering effect that is improving the fit estimate. The correlation and difference methods gave similar results on all of the images tested, and were not significantly affected by noise apart from on image E. This lack of sensitivity to noise results from the large area over which the correlation and difference is calculated (128x128 window). As expected, the correlation and difference methods do not perform as well on image E. This is because the approximation of the peak by a pyramid is no longer valid because the dominant feature size is less than 1 pixel. As the noise exceeds a critical level, the registration error jumps dramatically as the methods no longer reliably estimate the offset. Overall, the predictive registration method performed most consistently on all 5 images, with the results being almost identical each images. The phase and predictive methods consistently gave registrations to within 1-2% of a pixel at low noise levels, whereas the typical accuracy of the correlation and difference methods was 5-20%.

The effects of brightness and contrast were not investigated in this study because all of the low-resolution samples for a particular image were derived from a single high-resolution image. If the low-resolution images are captured by the same camera in quick succession, differences in the brightness and contrast between the low-resolution images should be insignificant. If these factors are significant, it is expected that the difference and prediction methods will not perform as well unless the images are pre-processed to remove or reduce the inconsistencies in brightness and contrast.

If the images are already registered to the nearest pixel, the computational complexity of the correlation and difference and prediction methods is similar. If necessary, a hierarchical approach could be used to efficiently give an initial registration to the nearest pixel. The phase method is significantly more computationally intensive, but does not require that the images be approximately registered first.

5 Conclusion

The correlation and difference methods perform poorly in the presence of significant fine (sub-pixel) detail because the pyramidal interpolation used to estimate the offset to sub-pixel accuracy no longer applies. The phase and predictive methods of registration give significantly better results, particularly in the presence of fine image detail. This makes these methods particularly suited to the registration of images for resolution improvement applications. Although the phase and predictive methods are more sensitive to noise than the more conventional methods, if the level of noise is such that the registration is inaccurate, then the noise is also likely to cause significant problems with the inverse filtering.

The predictive interpolation method involves significantly less computation than the phase method, making it the method of choice if the images are already registered to the nearest pixel. As it is a new method, it warrants further study as a general registration technique.

References

[1] R.C. Hardie, T.R. Tuinstra, J. Bognar, K.J. Bernard, and E. Armstrong: High resolution image reconstruction from digital video with global and non-global scene motion. *Proc. IEEE International Conference On Image Processing*, Vol I (October 1997) 153-156.

[2] M.C. Hong, M.G. Kang, A.K. Katsaggelos: An iterative weighted regularized algorithm for improving the resolution of video sequences. *Proc. IEEE International Conference On Image Processing*, Vol II (October 1997) 474-477.

[3] D.G. Bailey: Super-resolution of bar codes. *Proc SPIE Vol 3521, Machine Vision Systems for Inspection and Metrology VII* (November 1998), 204-213

[4] D.G. Bailey: Image Capture Modelling for High Resolution Reconstruction. *Proceedings of Image and Vision Computing NZ*, (November 1998), 38-43.

[5] A.K. Jain: *Fundamentals of Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey (1989) 400-407.

[6] E.O. Brigham: *The Fast Fourier Transform*. Prentice Hall, Englewood Cliffs, New Jersey (1974).