Multi-Variate Cross-Correlation and Image Matching

R. B. Fisher and P. Oliver
Department of Artificial Intelligence
University of Edinburgh

Abstract

This paper introduces the use of a multi-variate correlation function for region-based image matching and extends it to a modified cross-correlation function that works well when matching image areas are required have the same intensity contrast. It also shows that the multi-variate case is a straightforward generalisation of the monochrome image case. Experiments with both MRI and RGB colour imagery are shown, along with comparisons with the Euclidean, Manhatten and L_∞ matching metrics.

1 Introduction

With the increase in available computing power, both through fast microprocessors and also through special purpose VLSI and board-level products, vision researchers have been again investigating image region-based matching processes. For example, research projects have investigated area-based stereo [3, 8], Ugaritic character stroke location [2], general template matching [1, 5, 10], MRI image correspondence determination over time [11], corner detection [6] and face recognition. What characterises all of these examples is the use of small image windows from a first (or model) image as templates for matching neighbourhoods in a second image. For example, the stereo matching processes attempt to match many regions in the first image to corresponding points in the second image (as an alternative to feature-based matching).

Many alternative image match evaluation functions have been considered:

- $\sqrt{\sum(x_i - y_i)^2}$ (i.e. Euclidean metric),
- $\sum \mid x_i - y_i \mid$ (i.e. Manhatten or L_1 metric) and
- $max_i(\mid x_i - y_i \mid)$ (i.e. L_∞ metric)

where: i indexes over N paired signal samples x_i and y_i. In our opinion, the standard statistical correlation function (1) is more well-suited as the measure of image similarity, because, when attempting to match image regions, one must consider the likely distortions between the images being matched. If we assume only small rotations (a few degrees) and small non-Euclidean (e.g. shear or projection) distortions, then the geometry of a small region does not change appreciably between images. However, what can change easily is the translation (due to changes in camera position or target motion), and the intensity distribution. A standard model for the intensity differences between two images is:

$$B_i = \alpha A_j + \beta$$

where i and j index corresponding pixels. (A and B are the intensity levels of the pixels, α is the gain difference and β is the offset difference). Causes for this sort of linear intensity relationship might be: digitiser base level differences, digitiser gain differences, illumination intensity differences, shadows, changes in light direction (of distance light sources), etc. When images satisfy this model, then the three matching functions mentioned above do not perform well, as they assume that the images being matched have the same intensity distributions. However, the standard statistical correlation function can cope with this model.

The final thread in the introduction to this paper is the topic of multi-spectral images. With the increase in computing power, and the relatively recent commonplace availability of multi-spectral imagery (e.g. MRI proton density and difference, R/G/B video, multi-channel satellite remote sensing, registered range and intensity or reflectance data), one might now consider how best to match images whose pixels are vectors, rather than scalars.

2 Multi-Variate Cross-Correlation

From elementary statistics, the standard statistical cross-correlation function between two signals $\{x_i\}$ and $\{y_i\}$ is defined as:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \tag{1}$$

where: $\bar{x}$ and $\bar{y}$ are the means and σ_x and σ_y are the standard deviations of the N x_i and y_i. The function ρ takes values from [-1,+1], where values near +1 mean good correlation (i.e. when one function increases, the other also does in proportion), values near 0 mean uncorrelated (i.e. there is no relation between changes in one function and the other) and values near -1 mean anti-correlated (i.e. when one function increases, the other decreases in proportion).

In the case of matching image data, x_i and y_i are corresponding image pixels and i indexes over pixels in a neighbourhood geometry. As we are expecting to find regions that match, here the closer the value of ρ is to +1, the better the match. Note that subtracting the mean is useful in image matching, as it corrects for differences in digitisation base-level. Dividing by the standard-deviation is useful because it corrects for multiplicative effects, such as change in average illumination and digitisation contrast.

In the case of multi-dimensional image data, one might expect that this case is discussed in standard texts on multi-variate statistics. Indeed, there are correlation functions from one variable to vectors, and between components of vectors, holding other components fixed. However, an examination of over 30 texts on multi-variate statistics revealed only one page on vector correlation [7], which we believe was actually incorrectly formulated (see discussion below). Geiss *et. al.* [4] proposed a multi-variate cross-correlation function: Define the image samples to be $\vec{x}_i$ and $\vec{y}_i$, both of dimension M. As before, assume that there is a set of matched pixels $i = 1..N$ from some appropriate image neighbourhood. Then, the mean pixel values in the neighbourhood are:

$$\hat{\vec{x}} = \frac{1}{N}\sum_{i=1}^{N}\vec{x}_i$$

and similarly for $\hat{\vec{y}}$. The covariance matrix for each neighbourhood is:

$$\Lambda_x = \frac{1}{N-1}\sum_{i=1}^{N}(\vec{x}_i - \hat{\vec{x}})(\vec{x}_i - \hat{\vec{x}})^T$$

and similarly for Λ_y.

Now, Geiss *et. al.* define their correlation measure as the sum of the positive eigenvalues of the matrix:

$$S = \sum_{i=1}^{N}\vec{x}_i\vec{y}_i^T$$

This is not satisfactory, as (a) it is a measure that is sensitive to the absolute level and contrast of the data and (b) using only the positive eigenvalues does not allow reduction to the univariate case, and (c) using only the positive eigenvalues ignores the possibility of anti-correlation between the signals as might occur with repeated texture (i.e. a bivariate signal that had one perfectly correlated and one perfectly anticorrelated signal would be represented with an incorrect correlation). Therefore, we argue that one should use all of the eigenvalues.

To normalise the correlation for signal contrast and gain, the uni-variate case subtracts the mean and divides by the square-root, which suggests an obvious generalisation of subtracting the mean vector and pre-multiplying by the inverse of the square-root matrix:

$$Q_x^{-1}(\vec{x}_i - \hat{\vec{x}})$$

where

$$\Lambda_x = Q_x Q_x$$

However, by definition:

$$\mathbf{E}[Q_x^{-1}\vec{x}_i\vec{x}_i^T Q_x^{-T}] = I$$

(assuming $\hat{\vec{x}} = 0$ for simplicity). This pre-multiplication has the effect of "rotating" the data into a new coordinate system where there is no covariance between the components of the data. This leads to problems when considering the cross-correlation between two signals, each of which might have been "rotated" differently. This error also appears in another function for multivariate trace correlation discussed in [7]:

$$\frac{1}{M}trace(\Lambda_y^{-1}S^T\Lambda_x^{-1}S)$$

We argue that a better approach is to instead use

$$\Omega = \frac{1}{N}\sum_{i=1}^{N}D_x^{-1}(\vec{x}_i - \hat{\vec{x}})(\vec{y}_i - \hat{\vec{y}})^T D_y^{-1}$$

where D is a diagonal matrix with $(D_x)_{jj} = \sqrt{(\Lambda_x)_{jj}}$. That is, we normalise by the standard deviation of each component of the vector, after having subtracted the mean vector. Note that this has a similar form to (1). Then, the multi-variate cross-correlation coefficient is:

$$\rho = \frac{1}{M} trace(\Omega)$$

A little algebraic simplification of this sum reveals:

$$\rho = \frac{1}{M} \sum_{j=1}^{M} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{(x_{ij} - \hat{x}_j)}{(D_x)_{jj}} \frac{(y_{ij} - \hat{y}_j)}{(D_y)_{jj}} \right) \tag{2}$$

This is the mean of the cross-correlations between the individual channels.

3 Application to Image Matching

The most straight-forward application of the multi-variate correlation function described above is defined using a $S \times S$ window as the source of the N data samples. Then, for each appropriate window $\{\vec{x}_i\}$ in the first image, consider all appropriate windows $\{\vec{y}_i\}_k$ in the second image. (We use the term "appropriate" here because some aspect of the problem might limit the number of windows in the original image, such as sub-sampling or only using windows whose standard deviation was sufficiently large, or might limit the number of windows in the second image, such as only along an epipolar line, or in a bounded window.) For each window pairing, compute the correlation ρ_k and choose as the best match the window k that maximises ρ_k.

When using (2) for image matching, it is possible for image regions to match those that have very different intensities and contrasts, but whose local image intensity variations are very similar. If we have the constraint that contrast differences are small, i.e. if there are only small changes in scene illumination or signal output, to allow this contrast correction is unhelpful. So, an alternative contrast-constrained cross-correlation calculation is:

$$\rho = \frac{1}{M} \sum_{j=1}^{M} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{(x_{ij} - \hat{x}_j)(y_{ij} - \hat{y}_j)}{(max((D_x)_{jj}, (D_y)_{jj}))^2} \right) \tag{3}$$

If the two image regions are slight variations of the same scene, then this heuristic change has little effect. On the other hand, if there is a large difference in contrast between the matched regions, then the modified normalisation forces the correlation to zero (i.e. effectively uncorrelated).

Finally, contrast normalisation allows matching with regions that are essentially uniform except for random or texture variations. Our experience suggests that matching should not be allowed between a pair of regions if the sum of variances of either region is low. (This constraint is applied in the tests below).

4 Experiments

4.1 MRI Image Section Matching

The problem being investigated is how to identify and quantify plastic deformations of the brain over time (e.g. as arising from AIDS, Alzeheimer's disease, etc.) [9]. One part of that research required pixel-wise matching of image regions between two MRI brain image sections taken at different times (assuming that the same section had been selected for matching). The MRI scanner settings might be different, the patient might have moved, and the brain might have moved within the skull between measurement sessions. We assume that the brain has not rotated or changed significantly. Thus, the neighbourhood correlation method proposed above can be used. Figure 1 (left) shows a typical MRI PD cross-section, (middle) shows the PD image taken at a later time. As well as there being a difference in the position of the brain, there are also some changes in the brain itself. Figure 1 (right) shows the registered initial T2 image (i.e. we have a 2-vector of data at each pixel).

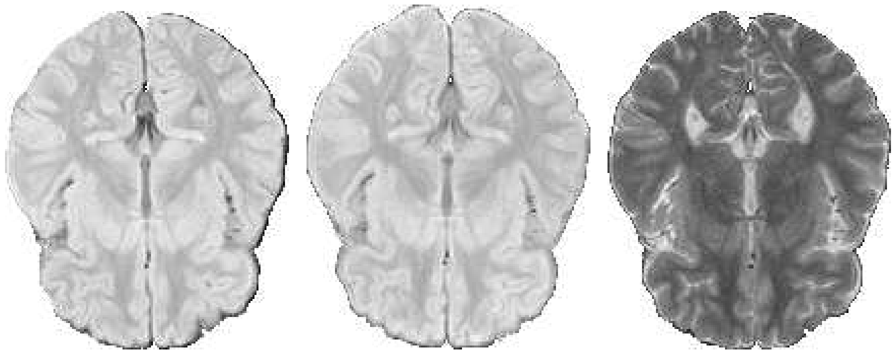

Figure 1: Measurements of a brain cross-section (left) PD data at the initial time and (middle) PD at a later time and (right) T2 data at the initial time.

Woodward [11] used the scalar correlation function to match a single channel MRI image. Figure 2 (left) show typical results of his process, with the vectors that map from pixels in the initial image to pixels in the second image. The pixel in the second image that has the maximum correlation is selected. The mapping at the left was based on the standard correlation measure applied to the T2 image and at the right metric (2) with the T2 and PD data is used. Because there is a slight rotation between the two images, there is a general "swirling" of vectors about a point to the left of the middle of the brain. While there are some differences between the two brain sections due to changes over time (especially in the upper left), the main thing to note is that there are a lot of pixels whose mapping is radically different to the mapping of their neighbours. This is observable as a cross-hatching effect in the mapping vectors. That is, the wrong pixel correspondences have been found. In (left), this mis-mapping is rather frequent; however, in (right), many fewer "cross-hatched" regions can be seen.

Figure 2: Mapping vectors between the initial and later images (left) using T2 correlation only and (right) using both T2 and PD data. A perfect match output should show all vectors "swirling" about a point just left of the centre of the brain. A mis-mapping is observable as a cross-hatching of overlapping vectors, whereas good mappings are largely all parallel.

The mappings were computed over a 139 by 166 set of sample points on a 50 MHz 486DX PC (with FPU) and took about 30 minutes for the T2 correlation only and 60 minutes for the correlation using both T2 and PD data. A 9×9 window is used for the correlation, and search is restricted to a 11×11 neighbourhood about the initial position. Figures 3 and 4 show the mapping functions on the PD/T2 vector pairs for the five matching functions described. The results seen here and in the table below show that the standard correlation metric given in (2) is substantially better than the other algorithms. In this case, there are few areas with constant contrast. A match was considered to be bad if it was more than 3 pixels from its expected position. For the 17,647 matches tried, the summary of matching is:

Metric	**Time**	**Bad**	**Mean Error Dist (Pixels)**	**Error Std Dev**
Correlate	60 m 05 s	806	0.58	1.28
Modified Correlation	59 m 19 s	1176	0.77	1.44
Euclidean	15 m 54 s	1583	0.92	1.63
Manhatten	14 m 18 s	1600	0.93	1.62
L_∞	14 m 37 s	1812	1.05	1.69

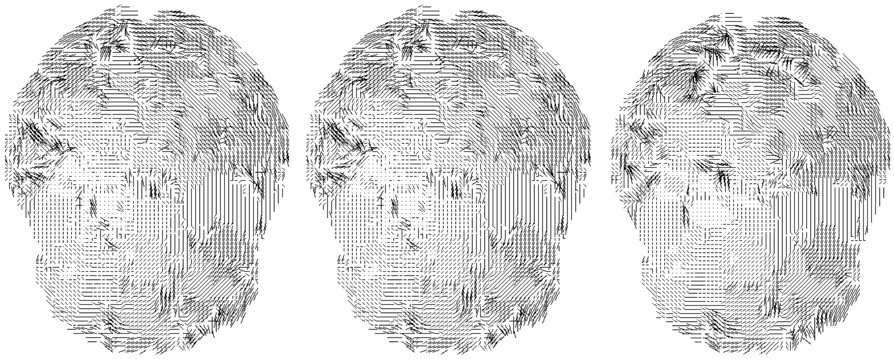

Figure 3: Mapping vectors using (left) the correlation metric given in (2), (middle) the contrast weighting correlation metric given in (3) and (right) the Euclidean metric.

4.2 RGB Colour Image Matching

We use the correlation function to derive a dense pixel based (as compared to feature-based) image correspondence between a stereo pair of colour R/G/B images. In this case, the image pixels are 3-vectors. Figure 5 shows the initial R/G/B image planes. The images were captured from a collage of three different photographs with different spectral characteristics and image textures. Here, the initial image is about 20 pixels to the left of the scene in the second image. As the images were taken at different times, there is a slight difference in camera positions, projection, illumination and noise between the two sets of images.

For each pixel in the left image, a 7×7 window about that pixel is used for matching when searching in the right image for the corresponding pixel. (We assumed that disparity limits the maximum shift of a pixel to ± 20 pixels about the expected position.) The pixel selected as the matching pixel is that with the maximum correlation, provided that

$$\frac{1}{M} trace(D_x) > \tau \qquad \text{and} \qquad \frac{1}{M} trace(D_y) > \tau$$

(here, $\tau = 7.5$. Using 12.5 reduced the number of matches allowed slightly, but had no significant effect.). The reason for this condition is that regions where there is little contrast (i.e. uniformly lit, uniform reflectance regions) correlate well with any other region also having little or no contrast, and this test eliminates regions whose contrast is low. Additional conditions may be needed for uniform intensity *gradient* regions (where any nearby pixel will have the same correlation).

Figure 6 shows the R channel and R/G/B multivariate mappings. As there was only a linear shift between the two views, all mapping vectors should be approximately horizontal. While the two mappings are quite similar, it is clear that the mappings computed using only the R channel (left) have more incorrect mappings than those computed using all of the R/G/B data (right). Analysis of the mappings shows that only 389 of 506 possible mappings were acceptable (by

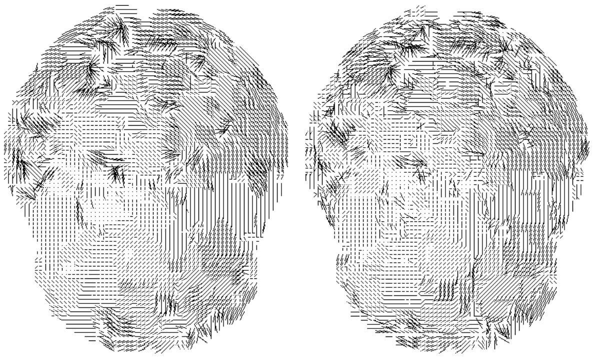

Figure 4: Mapping vectors using the metric given by (a) the Manhatten distance and (b) the largest absolute value.

Figure 5: (left) Red, (middle) Green and (right) Blue components of initial image

the average channel standard deviation being at least 7.5 in the window) with the R channel matching, of which 297 were good (i.e. mapped positions were less than 4 pixels from expected position). In the case of the RGB matching, 415 mappings were acceptable of which 321 were good. When the modified contrast condition was used, then 351 good matches were found. Thus, the multi-spectral matching gave more matched regions with fewer bad matches. We also compared the performance using the Euclidean, Manhatten and L_∞ metrics. In all of these cases, 415 mappings were acceptable using the same criteria as the RGB case, and the number of good matches was 357 for the Euclidean metric, 363 for the Manhatten metric and 350 L_∞ metric. Thus, in this case, the Manhatten metric has advantages over the other metrics; however, as the image distributions are the same between the two images, one would expect the non-normalising metrics to have better performance. None-the-less, all of the vector metrics were better than the scalar metric. The mappings were computed on a SparcStation 10/50 and required about 2 minute for the R channel and 6 minutes for the R/G/B mappings.

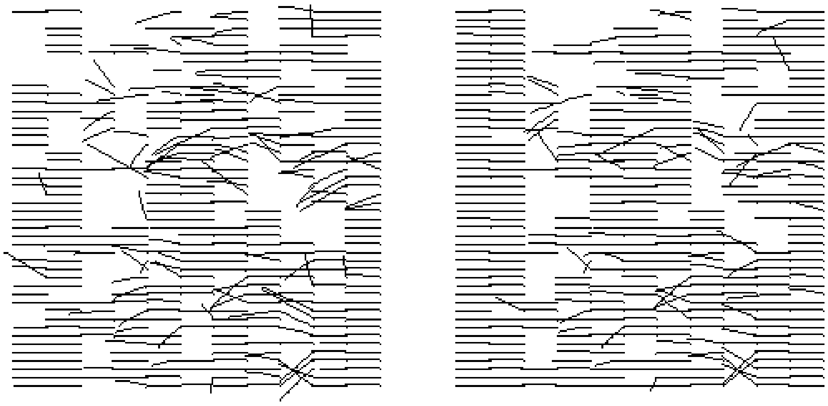

Figure 6: Mapping vectors between the initial and later images (left) using R channel correlation only and (right) using all R/G/B data with the modified contrast metric.

5 Conclusions

Theoretical analysis has shown that an effective multi-variate cross-correlation coefficient is simply the mean of the cross-correlation coefficient of the individual channels. The measure proposed here is also an improvement on Geiss [4] by accounting for normalisation of the difference signal channels and use of negative eigenvalues. This measure can be adapted slightly for image region matching, to eliminate matches using regions with low contrast. Experiments with many vectors in several images have shown that the use of the multi-spectral data does reduce the number of bad matches in actual use. The choice of metric to use depends on whether the image intensity distributions are identical between images being matched. If they are not identical, then the normalising effect of the correlation metrics is essential; otherwise, the non-normalising metrics (e.g. Manhatten or Euclidean) may have better performance as well as be faster.

The computational complexity of this process can be rather high in its worst form of complete image matching, as there is an N pixel correlation calculation for all S^2 possible correspondences between two images each of size S. Thus, this process will need special purpose hardware

This matching process is only suitable when the transformation between images is limited to translation of image positions and linear transformations of pixel values. When rotations are small (e.g. less than 5 degrees), the process will still work; however, for larger rotations the windows being matched no longer correlate well with the rotated version of the correct match. The same problem arises with scaling and shear of the image pixel positions. If this assumption does not hold, the correlation process is still valid; however, in this case, one would have to search for the maximum correlation with rotated and scaled versions of one of the images (e.g. every 5 degrees), unless some global estimation of rotation and scale was possible and then used to remove these effects. While this increases the computational costs, the principle remains the same.

Acknowledgements

The authors would like to thank the University of Edinburgh for support. Thanks also to D. Eggert and A. Fitzgibbon for advice and J. Best and E. Rimmington of Medical Radiology for the head data.

References

[1] J. Altmann and H.J.P. Reitbock. "A fast correlation method for scale- and translation-invariant pattern recognition". IEEE Trans PAMI, Vol 6, 46-57, 1984.

[2] T. Anthoni. "Recognition and Location of Ugaritic Character Stylus Strokes from Clay Tablet Images". MSc Dissertation, Dept. of Artificial Intelligence, Univ., of Edinburgh, 1994.

[3] L. Cohen, L. Vinet, P.T. Sander, and A. Gagalowicz. "Hierarchical region based stereo matching". Conf. on Computer Vision and Pattern Recognition, 416-421, 1989.

[4] S. Geiss, J. Einax, K. Danzer. "Multivariate correlation analysis and its application in environmental analysis". Analytica Chimica Acta, 242, 5–9, 1991.

[5] A. Goshtasby, S.H. Gage, and J.F. Bartholic. "A two- stage cross correlation approach to template matching". IEEE Trans PAMI, Vol 6, 374-378, 1984.

[6] X. Li, C. Shanmugamani, T. Wu, and R. Madhavan. "Correlation measures for corner detection". Conf. on Computer Vision and Pattern Recognition, 643-646, 1986.

[7] K. V. Mardia, J. T. Kent, J. M. Bibby Multivariate Analysis, Academic Press, London, 1974 (section 6.5.4, pp 170–171).

[8] S.B. Marapane and M.M. Trivedi. "Region-based stereo analysis for robotic applications". IEEE Trans Sys. Man and Comp, Vol 19, 1447-1464, 1989.

[9] P. I. Oliver. "Extensions to a Change Detection in MRI Brain Scan Data Program". Honours Dissertation, Department of Artificial Intelligence, University of Edinburgh, 1995.

[10] H.S. Ranganath and S.G. Shiva. "Correlation of adjacent pixels for multiple image registration". IEEE Trans Comp., Vol 34, 674-677, 1985.

[11] J. Woodward. "Change Detection in MRI Brain Scan Data". MSc Dissertation, Dept. of Artificial Intelligence, Univ., of Edinburgh, 1990.