7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho.

Enhancement of image-to-image co-registration accuracy using spectral matching methods

Keith Rennolls[1] and Mingliang Wang[1,2,3]

1 School of Computing and Mathematical Sciences, University of Greenwich
London SE10 9LS
k.rennolls@gre.ac.uk

2 Warnell School of Forest resources, UGA
USA.
mwang@smokey.forestry.uga.edu

3 Chinese Academy of Forestry, Beijing, PRC.

Abstract

Two of the important stages in the production of maps from remotely sensed imagery are rectification and classification. Residual geometric errors following rectification produce a "component" of the apparent classification error. Hence, besides being of value in its own right, reduction of the rectification error will enhance the quality of the classification stage, and the classification accuracy statistics will more closely refer to "pure" classification error rather than classification error due to pixel mismatch. In rectification to ground control points (GCPs) there is a natural limit of what can be achieved in rectification of an image, and this is determined by the cost of collecting good GCPs. However, if we are concerned with a sequence of images, and are concerned primarily with estimating change and growth, the rectification is image-to image, and the process becomes one of pixel matching, and the only cost is processing time. We present a spectrally-based pixel-matching algorithm which seems to offer considerable scope for very accurate pixel-to-pixel and hence image-to-image matching. An algorithm is used which maximizes local correlations in each spectral band at each co-registration point: a multivariate anisotropic spatial auto-correlation approach. The results are demonstrated and validated in a case-study of a 1987 Landsat5 TM image, 1997 Landsat5 TM image and a 2000Landsat7 ETM image, all of the same forested region in China.

Keywords: image-to-image, pixel-to-pixel, co-registration, spectral-correlation-search, sub-pixel accuracy

1 Introduction

Constructing a land-use map from a single remotely sensed image is a long and complex process and there are many sources of error that arise at various stages of the process, (Foody, 2000), including the confounded errors of rectification, re-sampling and classification, (Rennolls, 2002).

The remotely sensed data available for map making is generally much more complex than the use of a single image. Available information usually comprises:

- A time-sequence of images, (at varying time intervals), with each image having partial spatial coverage.
- The resolutions of the multiple images are generally different at different times and with the spectral information available also varying between images, due to evolving sensor technology.

- The images are also likely to exhibit differing geometric orientations and geometric distortions.

Integrated use of such information would involve the use of spatio-temporal modelling in order to achieve interpolation and prediction. Such "conflation" has long been on the research agenda, UCGIS(2004), but the general challenge still remains (IMGFSR 2004a, Le Moigne 2004, NASA 2004b). Joyce and Olsen (2000) made an interesting attempt at integrated classification and growth modelling on multi-temporal imagery, based on a "radiometric correction" method which used a measurement errors model which provided a symmetric correction method between successive images. Rennolls (2004) argues that such complex historical imagery data is the key to future prospects for remote-sensing based estimation of change and growth.

There is a very full and technical history and literature in the area of registration and geometric correction. See Brown (1992) and Zitova and Flusser (2003) give reviews. Mahdi and Farag (2002) use a pair of Genetic Algorithms to match images; Mostfa et.al(2004) use Neural Networks; Cole-Rhodes et.al.(2003) use Mutual Information Maximization using a stochastic gradient optimization techniques; Wang et.al. (2004) compares cosimulation, and collocated cokriging methods with regression methods; Dowman and Dare 1999) match Polygons; Wilson and Sader (2002) use NDVI and NDMI. There is much related work in medical imaging, eg. Bardera et.al. (2004). Jiancho and Chern(2001) is particularly interesting in the way that combines the estimation of the registration transformation with the estimation of a shading function, with a robust weighted least squared method, where the weighting eliminates changed pixels from figuring in the estimation.

1.1 The importance of perfect image co-registration

A production map in which the rectification has not been perfect, (and this is always the case), remains geometrically distorted. Residual geometric errors in a production map produce a "component" of the apparent classification error, Rennolls (2002). Hence in an ideal world it is of utmost importance to reduce the rectification error as much as possible. The quality of the classification stage will then be enhanced, and the classification accuracy statistics will really refer to "pure" classification error rather than classification error due to pixel mismatch.

In rectification to GCPs there is a natural limit of what can be achieved in rectification of an image, and this is determined by the cost of collecting good GCPs. However, if we are concerned with a sequence of images, and are concerned primarily with estimating change and growth, the rectification is image-to image, and the process becomes one of pixel matching. The only cost is processing time.

In this paper we go into some detail on the pixel-matching problem, and present a simple heuristic spectrally based pixel matching algorithm which seems to offer considerable scope for very accurate pixel to pixel image matching.

2 Case Study Data

Three Landsat Images were provided by the Chinese Academy of Forestry. They are 1987 Landsat5 TM, 1997 Landsat5 TM, Landsat7 ETM. Each has previously been separately subjected to correction, both geometrically and radiometricly. The 1987 and 1997 images are in the TM/Krasovsky projection with pixel size of 35m, while the 2000 image in UTM/WGS 84 with pixel size of 33m.

3 Co-registration: Image to Image matching

The procedure suggested has two stages. The first stage is a standard co-registration of images based on a set of corresponding points on the images. This stage is only meant to provide the first approximation to a matching of images, and for convenience the co-registration points are chosen by eye. Stage 2 makes use of spatial correlation measures in multiple spectral bands to obtain a significant increase in the accuracy of the image co-registration.

3.1 Stage 1 Matching

If the matching of two or more images is to be done completely automatically, the problem is "hard" (computationally NP complete; Keysers and Unger, 2003). The Rectification/ Geometric-Correction approach to matching is usually "spatially" based. Matching control points have to be selected on each of the images and their coordinates determined as accurately as possible. If this determination is without error then a rubber-sheet transformation, such as a thin plate spline interpolates the match across the whole image. In such an approach the RMS is zero and the matching accuracy (or equivalently the matching model-error) has to be determined from a validation set of control points that have been reserved from the rectification process. Le Moigne (2004), NASA(2004b) are used Wavelets to get first stage co-orientation between images at a low resolution. Jianchao and Chern (2001) used a hierarchical segmentation and matching algorithm for such first stage orientation matching.

Ironically this image matching is largely trivial for the human eye-brain complex. From observation of the images in Figure 1, we immediately "see" the lower "trident foot" of the common "snakes" in our images, whether they be ridges, valleys or roads. Our mind's-eye immediately (i) matches the images with just with a handful of corresponding features, and mentally (ii) superimposes the intervening regions on the images. We bi-pass the hard problem of whole image matching, and follow the human eye-mind heuristic through two phases, as indicated above. First we use the (actual) eye to select some corresponding control points, and roughly determine their coordinates in their own image frames. As few as two points would be sufficient for the matching using a linear mapping, three for a projective mapping, and four for a perspective mapping, (Glasbey and Mardia 1998).

3.1.1 Selecting Control Points (CPs)

If the matching For the 1987 and 1997 TM images, 41 CPs were selected manually. They are mainly road or river intersection points, which are relatively easily recognized in images. Figure 1 shows some of the stage-1 control points, mostly road/road or road/river intersections

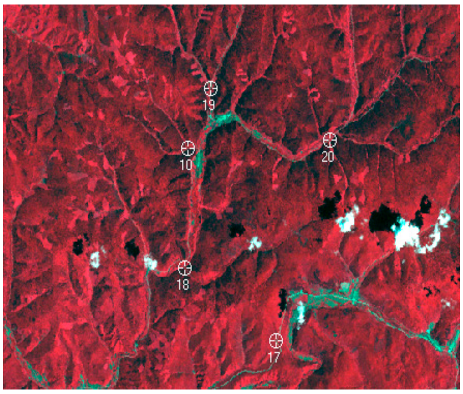


Figure 1 Several manually selected GCPs in the 1997 image for stage-1 matching.

3.1.2 The 1987-1997 Perspective Transformation

The general perspective transformation is given by:

$$u = \frac{a_0 + a_1 x + a_2 y}{1 + c_1 x + c_2 y} \qquad v = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y} \tag{1}$$

where (u, v) are 1997 image coordinates, and (x, y) are the 1987 image coordinates (IMGFSR 2004b, Glasbey and Mardia 1998). Table 1 gives the estimated parameters, using ordinary least squares (OLS) to fit (1) to the 41 control points.

Note that c1 and c1 are essentially zero, so an affine/linear transformation is sufficient. Note also that for the affine transformation, we essentially have a1 = 1 and a2 = 0. Similarly we have b1=0 and b2 = 1. Hence the (x,y) and (u,v) already have the same orientation (no rotation is needed) and the same scale! The transformation reduces to a simple translation of the origin, given by the (a_0, b_0) vector.We should have known this, but did not**!** We suspect that this fact is not as well known as it should be by those using Landsat imagery.

Table 1 OLS perspective transformation 1987→1997, on 41 “rough” points.

Parameter	From 41 CP's
a0=	10882.83622
a1=	0.975338626
a2=	-0.002058339
b0=	112205.3522
b1=	-0.005667355
b2=	0.953838592
c1=	-1.29E-09
c2=	-4.75E-09

3.2 Stage 2: spectral pixel matching for refinement of image-to image co-registration

This stage of co-registration refinement uses within spectral band intensity information, and is illustrated in Figure 2. Let A be a pixel in image 1, on the left-hand-side of Figure 2.

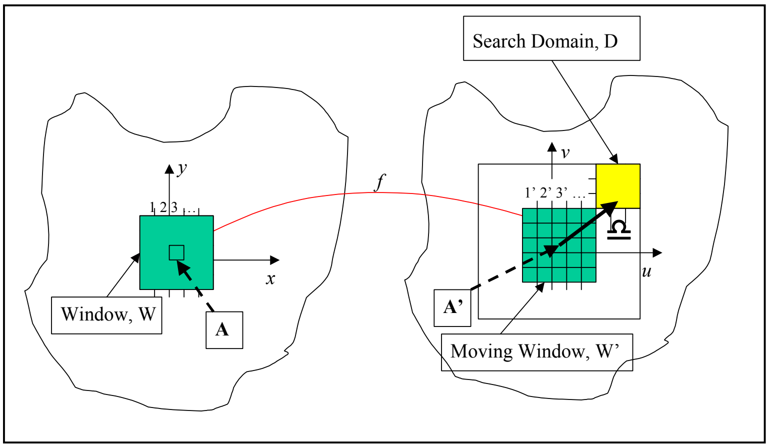


Figure 2 Spectral refinement of pixel-to-pixel matching.

Let A'=*f*(A) be the corresponding pixel in image 2, under the Stage 1 image-to-image co-registration transformation *f*. Let $\delta = (\delta_u , \delta_v)$ be an off-set from A', where δ_u and δ_v are the off-sets from A' in pixels in the directions u and v, the directions of the coordinate axes of image 2. Denote the pixel in Image 2 which is δ from A' by B(δ). If δ is small, then B(δ) is a pixel in the neighbourhood of the image of A, in image 2. We define a square neighbourhood of pixels around A' which we call the Search Domain, Δ say. We are interested in how "similar" the environments of A and B(δ) are. We define a equal size square Sensing Windows (of size mxm) around A and B(δ), in images 1 and 2 respectively, denoted by W and W'(δ) respectively. In each of these two sensing windows we label the pixels in the same sequence; 1, … , M (= mxm) (relative to the reference axes in each image). We then use the correlation between the Band-I intensities in the pixel-sequences 1, …, m in each image as the measure of local similarity, (I= 1, 2, …7). These measures may be regarded as cross-image (structured) spatial correlation functions. The "structured" term is used to distinguish from the more-oft used cross-(|δ|-dependent)-correlation function.

We search over δ in Δ(A') to find the δ such that Corr(W_I(A), W'(B(δ)) is maximum. We take this point B(δ), in image 2, to be the "spectrally refined" matching point of A in image 1. This is an intuitively obvious local correlation search algorithm, which is similar to the use correlation search described by Schowengerdt (1997).

We have tried other spectral distance measures, but the correlation measure has been best. It is fairly clear, intuitively that the correlation as defined here, will be more able to pick up on the rather unique nature of the control points which have been selected. [a CP would never be simply selected in the middle of a homogenious region]. It is therefore likely to be superior to the Information Divergence measure (MI) (which depends on an assumption of independent

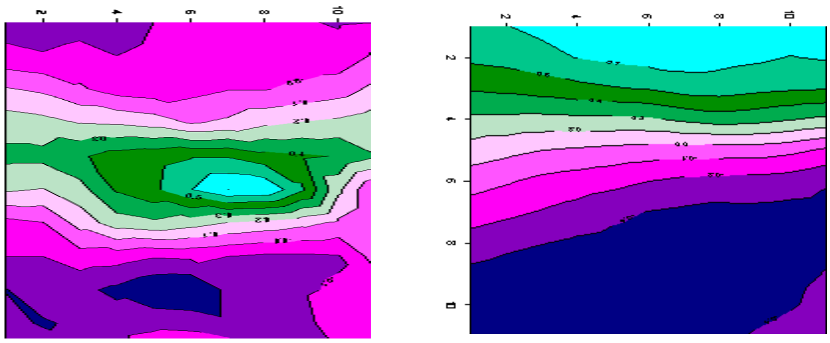
Multinomial pixel type), or a spatial autocorrelation measure, (or kriging equivalents) (which assume isotropy in the image, and force it into the measure).

When calculating these cross-correlational measures we have evaluated the use of 3*3, 5*5, 7*7, 9*9, and 11*11 sensing windows. However, in this paper we report only on the use of an 11*11 search domain, and an 11*11 sensing window, (the sensing window may extend beyond the search window).

	1	2	3	4	5	6	7	8	9	10	11
1	-0.18	-0.2	-0.18	-0.15	-0.13	-0.1	-0.13	-0.13	-0.12	-0.14	-0.13
2	-0.1	-0.12	-0.09	-0.13	-0.12	-0.09	-0.11	-0.1	-0.1	-0.09	-0.01
3	0.09	0.06	-0.01	-0.07	-0.09	-0.06	-0.03	-0.03	0	0.03	0.11
4	0.24	0.22	0.18	0.17	0.14	0.07	0.1	0.16	0.19	0.21	0.25
5	0.32	0.32	0.38	0.44	0.46	0.49	0.49	0.43	0.41	0.39	0.3
6	0.2	0.18	0.25	0.41	0.48	0.6	**0.71**	**0.7**	0.52	0.21	-0.03
7	-0.06	-0.03	0.12	0.24	0.27	0.23	0.18	0.15	0.11	0.06	-0.03
8	-0.13	-0.13	-0.07	-0.08	-0.12	-0.11	-0.13	-0.15	-0.15	-0.14	0.02
9	-0.24	-0.16	-0.21	-0.26	-0.33	-0.34	-0.21	-0.15	-0.14	-0.12	-0.1
10	-0.19	-0.25	-0.21	-0.18	-0.16	-0.17	-0.21	-0.19	-0.16	-0.08	-0.05
11	-0.06	-0.13	-0.18	-0.16	-0.04	0	-0.02	-0.11	-0.19	-0.14	-0.02

Figure 3 11x11 Correlation Map. Control Point 1, Band 1, using an 11x11 search window.

Figure 3 shows the calculated correlation table for the 11 *11 search domain for CP1, between 1987 and 1997. The A' pixel is are the centre, where the blue row and column intersect. We see the two highest values of the local correlation measure are to the immediate right of A'. We conclude there is (**about**) one pixel error in the original matching of CP1. Figure 4 shows a contour map of the local correlation structure constructed using S-Plus.



(a) Band 1. (b) Band 6 (Thermal).

Figure 4 Correlation Map: Control Point 1.

Figure 4(a) suggests that by fitting a suitable response surface model to this correlational map, it would be possible to estimate the pixel-to-pixel match with considerable sub-pixel resolution. Figure 4(b) shows that the thermal band 6 does not perform well in this approach, and should not be used.

Figure 6 shows the correlational maps for each of the bands (excluding the thermal band 6) for CP1 in the 1987-1997 matching exercise. We see that we have substantial mutual support in each of the bands for determination of the improved pixel to pixel match. It seems that in such a case, individual pixel matching should be possible to within about one ten of a pixel, (i.e. to an accuracy of within 3 metres!). However, we do not pusue the quantification of theis sub-pixel accuracy here.

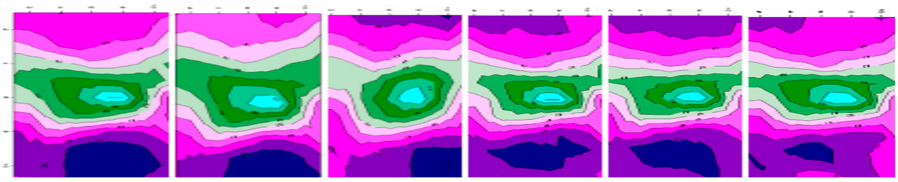


Figure 6 Correlation Maps: Control Point 1: All Bands (excluding Thermal).

3.4 A Multi-Spectral Pixel Matching Criterion

Let (x, y) be one image coordinates, (u, v) the other. For each band of 6 bands (b1, b2, b3, b4, b5, b7) of TM images, we got 6 "matching" results using CC as similarity measure for each CP. (u1,u2,u3,u4,u5,u7) for u corresponding to x, (v1,v2,v3,v4,v5,v7) for v corresponding to y.

So with partially matching (u1,u2,u3,u4,u5,u7), (v1,v2,v3,v4,v5,v7), for a given (x, y), how to determine our final selection of a "refined (u, v) matching point to (x, y)? We adopt:

Criterion 1: if more than half of (u1,u2,u3,u4,u5,u7) are within the same pixel, **and** more than half of (v1,v2,v3,v4,v5,v7) are in the same pixel, then the shared "u" and "v" are taken as the refined results for (u, v), expressed as (ur, vr).

For (x, y) points which do not satisfy this criterion we say the refinement has failed. The criterion chosen may be more or less strict than the one chosen and this will result in less or more points satisfying the refinement criterion.

We also do the reverse refinement, on (ur, vr) to give (xr, yr) (if it exists). We have :

$$(x, y) \rightarrow (ur, vr) \rightarrow (xr,yr)$$

Finally, if: Criterion 2: $(xr,yr) \equiv (x, y)$

is satisfied, then we have two-way matching and finally accept that we have a "Good Match Point" (GMP). This "round-trip" matching criterion is very strict and results in 20 GMP from the original 41 rough matching points for the 1987-1997 correspondence. There is clearly a trade-off between the strictness of the matching criterion and the adequacy of the final number of GMP points for the final image to image mapping determination. The optimum has not yet been determined.

4 Results for 1987 and 1997 Images

The rigorous multi-spectral criterion defined in the last section resulted in 20 GMPs being found from the original 41 “rough” point-pairs for the (87↔97) image matching. these matches are so accurate they should make use of a rubber-sheeting representation, such as thin plate splines to represent and interpolate them. However, for convenience in comparing with the previous perspective transformation of the 41 original rough control points the fit of a similar transformation is given in Table 2. We see that the a1 and a2 parameters are considerably closer to unity and zero than they were in Table 1, a consequence of the much more accurate spatial co-registration of the GCP’s than the original CP’s.

Table 2 Parameter estimates for the perspective transformation for the 20 GMPs (87↔97).

Parameter	From 20 GCP's	From 41 CP's
a0=	-6649.383485	10882.83622
a1=	1.005822864	0.975338626
a2=	0.002181113	-0.002058339
b0=	-53477.76408	112205.3522
b1=	-0.043547679	-0.005667355
b2=	1.028246821	0.953838592
c1=	-9.06E-09	-1.29E-09
c2=	3.55E-09	-4.75E-09

References

Bardera (2004) Normalized similarity measures for medical image registration.

http://ima.udg.es/iiia/GGG/UsersDocs/feixas/SPIE04_final.pdf

Bretschneider, T., Bones P.J., McNeill S. Resolution Enhancement Using Multispectral Remotely Sensed Imagery. http://www.ntu.edu.sg/home/astimo/Publications/Documents/IVCNZ1999.pdf

Brown L.G. 1992. A Survey of Image Registration Techniques. ACM Computing Surveys, 24(4):325-376.

Campbell, Norm, Furby, Suzanne Fergusson, Brian, “Calibrating Images from Different Dates”, http://www.cmis.csiro.au/rsm/research/calibration/calibrate.html.

Civco, D.L. and Waug, Y. (1994) Classification Of Multispectral, Multitemporal, Multisource Spatial Data Using Artificial Neural Networks. ASPRS/ACSM.

http://wwwsgi.ursus.maine.edu/gisweb/spatdb/acsm/ac94014.html

Cole-Rhodes, A.A., Johnson, K.L, LeMoigne, J, and Zavorin, I. (2003). Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Transactions on Image Processing, 12(12):1495-1511.

Dowman, I. and Dare, P. (1999) Automated procedures for Multisensor registration and ortho-rectification of satellite images. *International Archives of Photogrammetry and remote Sensing,* Vol.1, 32, Part 7-4-3.

Dundar, M. M. and Landgrebe, D.A. (2004) Toward an Optimal Supervised Classifier for the Analysis of Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*. Volume 42, No. 1, pp 271-277, January, 2004. http://dynamo.ecn.purdue.edu/~landgreb/toward_optimal.pdf

Foody, G.M. (2000) Accuracy of Thematic Maps Derived from Remote Sensing. pp217-224 in Heuvelink and Lemmens, 2000.

Fougnie B., Deschamps P.Y., Frouin R. (1999) Vicarious calibration of the POLDER ocean color spectral bands using in-situ measurements. *IEEE Transactions on Geoscience and Remote Sensing*, Volume 37, Number 03, p. 1567.

Franz, Bryan, “OCTS, MOS, and POLDER Vicarious Calibration Analyses”, SIMBIOS Science Team Meeting 13 September 1999, http://simbios.gsfc.nasa.gov/Documents/TeamMeetings/1999/Franz_OCTS_

POLDER_SIMBIOS99.pdf

Glasbey, C. and Mardia, K.V. (1998). A review of image warping methods. Journal of Applied Statistics, 25, 155-171. http://www.bioss.sari.ac.uk/image/warp.pdf

Goshtasby A. and Stockman G.C. (1986) A region-based approach to digital image registration with subpixel accuracy. IEEE Transactions on Geoscience and Remote Sensing, GE-24(3):390-399.

Hall, F.G., Strebel D.E., Nickeson, J.E. and Goetz, S.J. (1991) Radiometric rectification: Toward a common radiometric response among multidate, multisensor images. Remote Sensing of Environment, 35(1):11-27

Hayes & Sader (2004) Change Detection techniques for Monitoring Forest Clearing and Regrowth in a Tropical Moist Forest. http://www.ghcc.msfc.nasa.gov/corredor/change_detection.pdf

Igbokwe, J.I. (1999) Geometrical processing of multi-sensoral multi-temporal satellite images for change detection studies. International J. Remote Sensing, 20(6):1141-1148.

IMGFSR(2004a) Image Fusion Systems Research. http://www.imgfsr.com/

IMGFSR(2004b) Transformation Functions for Image Registration. http://www.imgfsr.com/ifsr_tf.pdf

Jianchao, Y. and Chern, C.T. (2001) The Practice of Automatic satellite Image registration. Presented at the 22nd Asian Conference on Remote Sensing, 5-9 Nov. 2001, Singapore.

http://www.crisp.nus.edu.sg/~acrs2001/pdf/072yao.pdf

Joyce, S. and Olssen, H., (2000). Long-term forecast monitoring with temporal-spectral trajectories from Landsat TM data. pp477-489 in Zawila-Niedzwiecki and Brack (eds), 2000.

http://rogow99.sggw.waw.pl/09_session_5/04/

Kettig R. L. and Landgrebe, D. A. (1976) Computer Classification of Remotely Sensed Multispectral Image Data by Extraction and Classification of Homogeneous Objects. *IEEE Transactions on Geoscience Electronics*, Volume GE-14, No. 1, pp. 19-26, January 1976.

Keysers, K. and Unger, W. (2002) Elastic image matching is NP-complete. Pattern Recognition Letters 24, 445-453.

Landgrebe (2004) Selected Publications. http://dynamo.ecn.purdue.edu/~landgreb/publications.html

Le Moigne, J. (2004) Image registration and Fusion for Future Formation Flying Systems.

http://is.arc.nasa.gov/IDU/slides/ImgFsn02c.pdf.

Li H. and Manjunath, B.. (1995) A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, 4(3):320-334.

Mahdi H. and Farag, A.A. (2002) Image registration in Mulltispectral Data Sets. IEEE ICIP II 369-372.

Mostfa, G. H. (2004) Multimodality Image registration and Fusion Using Neural Network.

http://www.cvip.uofl.edu/publications/cv/MostafaGHMostafa217.pdf.

NASA(2004a) Full Spectral Imaging - http://carstad.gsfc.nasa.gov/Topics/JBRESEARCH/FullSpect.htm

NASA(2004b) Image Registration and Fusion for Future Formation Flying Systems.

http://is.arc.nasa.gov/IDU/tasks/ImgFsn.html.

Rennolls, K., 2000. Algorithms and software for Landsat image classification with a case study on a tropical rain forest in Sumatra. pp523-536 in Zawila-Niedzwiecki and Brack (eds), 2000.

Rennolls, K. (2002) Components of Error Of Maps From Remotely Sensed Images. Spatial Accuracy 2002. http://rogow99.sggw.waw.pl/09_session_5/08/.

Rennolls, K. (2004) Satellite “Image-Banks” in Forest Information Systems? Prospects for Accurate Estimation of Forest Change and Growth from Data-Fusion Models. Spatial Accuracy Symposium, July 2004. Portland Maine, USA.

Schott, J., Salvaggio, C.and W. Volchok, (1988). "Radiometric Scene Normalization Using Pseudo-invariant Features", *Remote Sensing of Environment,* 26:1-16.

Schowengerdt, R. A. (1997) *Remote Sensing: Models and methods for Image Processing*. 2nd Edn. AP.

Secker, J., Staenz, K., Gauthier, R.P., Budkewitsch, P (2001) Vicarious calibration of airborne hyperspectral sensors in operational environments. *Remote Sensing of Environment*, 76, No. 1, pp. 81-92.

Teillet, P. M., Fedosejevs, G., Gauthier, R. P., O'Neill, N. T., Thome, Kurtis J., Biggar, Stuart F., Ripley, H., Meygret, A. (2001) A Generalized Approach to the Vicarious Calibration of Multiple Earth Observation Sensors in Hyperspectral Data. *Remote Sensing of Environment*, 77:3, pp. 304-327.

Wilson, E.H., and Sader, S.A. (2002) Detection of forest harvest type using multiple dates of landsat TM imagery. *Remote Sensing of Environment*, 80, 385-396.

Zawila-Niedzwiecki, T., and Brack E. (eds) (2000) *Remote Sensing and Forest Monitoring*; Proceedings of a IUFRO Conference, Rogow, Poland, 1999. Warsaw University Press. http://rogow99.sggw.waw.pl/

Zitova B. and Flusser J. (2003) Image registration methods: a survey. *Image and Vision Computing*, 21:977-1000.