Accuracy improvement and evaluation measures for registration of multisensor remote sensing imagery

Deniz Gerçek*, Davut Çeşmeci, Mehmet Kemal Güllü, Alp Ertürk, Sarp Ertürk
Kocaeli University Laboratory of Image and Signal Processing (KULIS), Electronics and Telecommunication Engineering Department, Umuttepe Campus, Kocaeli, Turkey

Abstract

Intensity based image registration methods are widely used in fine geometric registration of multisensor images. Accordingly, for images that are compared through translation of image templates, position where similarity measure is maximized is assumed to indicate best registration. Image registration quality is of crucial importance especially for studies that have high geometric accuracy requirements; e.g. image fusion, change detection, multichannel segmentation, and Digital Terrain Model (DTM) generation. Accuracy of image registration is conventionally evaluated by means of error measures (e.g. RMSE) obtained through comparison of coordinates of control points from the target and the reference / ground truth. However, especially for multisensor images with low spatial resolution component, difficulty in precisely positioning control points inhibits both sub pixel accuracy and evaluation of the registration. In this study, three widespread measures in intensity-based image registration namely, Normalized Cross Correlation (NCC), Mutual Information (MI), and Phase Correlation (PC) are tested for registering images acquired from EO-1 Hyperion and IKONOS sensors. We propose the use of ‘global similarity’ and ‘inverse consistency’ measures for evaluating the performance of these intensity based automated registration methods.

Keywords
Image registration, Intensity-based, Similarity measure, Accuracy assessment, Hyperion, IKONOS.

Özet

Farklı alıcıldardan elde edilen uydu görüntülerinin çıkarılmasında başarım artırma ve değerlendirme ölçümleri


Anahtar Sözcükler
Imge çakıştırma, Parlatıklık temelli, Benzerlik ölçütü, Başarım değerlendirme, Hyperion, IKONOS.

*Corresponding Author: Tel: +90 (262) 3033392 Fax: +90 (262) 3033003 E-mail: deniz.gercek@kocaeli.edu.tr (Gerçek D.), davutsmc@gmail.com (Çeşmeci D.), kemalg@ kocaeli.edu.tr (Güllü M.K.), erturk.alp@gmail.com (Ertürk A.), sertur@kocaeli.edu.tr (Ertürk S.),
### 1. Introduction

Multi-sensor image analysis has been a major field of interest in remote sensing. Bringing together images from different sensors, views and time periods is essential for extracting relevant information for particular studies. However, multi-modalities/multi-sensor characteristics are usually a challenge in bringing images into a common basis of comparison or aggregation. Registration of images in this context is a fundamental prerequisite. Along with that, particular studies e.g. data/image fusion (Mitianoudis and Stathaki 2007; Yokoya et al. 2012), resolution enhancement (Eismann and Hardie 2008), change detection (Dai and Khorram 1998), stereo surface modeling (Baltsavias et al. 2008) demand higher geometric accuracy standards that may necessitate quantitative evaluations. For such studies, accuracy errors not exceeding 0.1–0.2 of the low-resolution pixel size is expected for image co-registration (Dai and Khorram 1998; Zhukov et al. 1999).

Accuracy of image registration is conventionally evaluated by measuring error between control points that both exist in target and the reference images. However especially for the remote sensing images of very different spatial resolution it is barely possible to precisely locate corresponding ground control points both in high and low resolution images. Hence, there is no basic standard for evaluating accuracy.

Image registration, so-called image co-registration is defined as geometrically aligning images from different sources, different periods or views (Brown 1992). Registering images from different sensors has been a hot research topic in the literature for over a decade with the advent of sensors with various characteristics. There are numerous remote sensing satellites in mission. GeoEye and DigitalGlobe are the main commercial satellite organizations that provide high spatial resolution multispectral images. Hyperspectral images on the other hand provide high spectral resolution with contiguous and narrow bands in a wide range spectrum. Therefore, they outperform on target detection and classification over multispectral images. However, hyperspectral images have the disadvantage of low spatial resolution due to technical constraints that seem to last for the near future (Thomas et al. 2008). This encourages the research on enhancement of spatial resolution of Hyperspectral images. Hyperspectral and optical multisensor images are also complementary in a way that hyperspectral images are of high spectral and low spatial resolution, and multispectral images are vice versa. Therefore, an information content of high spectral-spatial resolution is anticipated from multisensor image analysis, provided that these two imagerys are registered with high accuracy.

In this study, image registration with high accuracy is proposed prior to analysis intended for data fusion of Hyperion and IKONOS multispectral imagery. Amount of information that can be gathered from multiple sources is significantly higher when compared to that of datasets would individually produce (Bunting et al. 2010). However there are also challenges in registering multiple-source remotely sensed data, especially when there is big difference in their spatial resolution.

In image fusion, ratio between spatial resolution of the high-resolution image and that of the low-resolution multispectral image; so-called ‘spatial resolution ratio’ is known to influence the quality of the fused image. Experimental results indicate that a spatial resolution ratio of higher than 1/10 is usually desired for optimal multisensor image fusion (Ling et al. 2008).

In image fusion approaches that intend information extraction or classification; high resolution content can be embedded into low resolution using spatial domain techniques. These techniques transfer high-resolution information from the high-resolution image to low resolution spectral bands using various deterministic or statistical predictors (Zhukov et al. 1999; Yokoya et al. 2012). High resolution content can also be utilized in the form of objects constructed from high spatial resolution image using object-based techniques (Mitianoudis and Stathaki 2007).

Image registration approaches are broadly grouped into two as feature-based and intensity-based (Zitova and Flusser 2003). Feature-based methods require a set of control points that can be detected both in the reference and the target image. As this is not quite possible for the images with very different spatial resolution, we have focused on intensity based registration methods that utilize intensity of image templates to match the image pairs.

Accuracy of each method is evaluated by means of ‘global similarity’ and ‘inverse consistency’ of transformation. High global similarity between transformed target image and the reference image is assumed to be an indicator of high registration accuracy. Similarly, higher consistency between forward and backward transformation is assumed to produce more accurate results.

### 2. Image registration methods and evaluation measures

Image registration has many potential applications in various fields of research including computer vision, medical imaging and remote sensing. During the last decades, growing amount and diversity of obtained images invoked the research on automated image registration. A complete view of image registration research in two decades and mainstream approaches can be found in Brown (1992), Zitova and Flusser (2003), Wyawahare et al. (2009), and Gruen (2012).

Image registration approaches are broadly grouped into two as feature-based and intensity-based (Zitova and Flusser 2003). Feature-based methods are applied through either manual or automated (e.g. SIFT, Harris) detection of control points that tie common features in images from different sensors. Intensity-based approaches rather make use of a similarity measure between the reference and the target and adjust the transformation until the similarity is maximized based on the assumption that the images will be most similar at the correct registration.

Intensity-based methods are effective in correcting translation and rotation errors in images; hence they are widely used for medical image registration where rigid transformation is usually satisfactory (Pluim et al. 2003). However, in the case of registering satellite imageries that are usually subject to local geometric distortions, intensity-based approaches tend to fail. To overcome this problem, a widely used technique is to shift area/blocks of target image across the reference image to calculate similarity between target
and the reference image templates that is called area-based approach (Ingладa and Giros 2004). Similarity measures calculate the similarity of templates of target and reference images, and translate target within a buffer zone across the reference to find the translation where they best match. These translations are then utilized to transform target image to match the reference.

Three similarity measures namely Normalized cross-correlation (NCC), Mutual Information (MI), and Phase Correlation (PC) that are widely used in the literature to register images are examined and tested for this study.

2.1 Similarity measures used in intensity-based image registration

Normalized Cross-Correlation (NCC) among similarity measures is widely used for registering images of the same modality. Correlation coefficient \( r \) as formulated in Equation 1 is assumed to be at its maximum for the best match.

\[
\begin{align*}
    r &= \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \left( (B_1(i,j) - \mu_1) \times (B_2(i,j) - \mu_2) \right)}{\sqrt{(B_1(i,j) - \mu_1)^2} \sqrt{(B_2(i,j) - \mu_2)^2}} \\
    \mu &= \text{mean value of } \text{describes the probability of a given intensity value}
\end{align*}
\] (1)

\( r \) is the correlation coefficient and \( \mu \) is the mean value of related image block ‘\( B \)’.

Mutual Information (MI) originating from ‘information theory’ is a measure of statistical dependency rather than linear relationship between image pairs as in NCC. Therefore it is particularly suitable for registration of multi-modal/sensor images (Fookes and Bennamoun 2002). MI calculates dependency as the difference between marginal entropies and joint entropy as shown in Equation 2 and is assumed to get its maximum when the images totally match. Normalized MI (NMI) calculated as in Equation 3 is reported to have improved on registration results compared to MI (Suganya et al. 2010).

\[
\begin{align*}
    MI(B_1, B_2) &= H(B_1) + H(B_2) - H(B_1, B_2) \\
    MI(B_1, B_2) &= H((B_1) + H(B_2)) / H(B_1, B_2) \\
    B_1(i,j) &= B_2(i-t_1, j-t_2) \\
    S[i,j] &= F^{-1} \left[ \frac{X_1[k] \times X_1[k] \times e^{-j\omega t} e^{-j\omega t'}}{X_1[k] \times X_1[k] \times e^{-j\omega t}} \right] = \delta[i-t_1, j-t_2]
\end{align*}
\] (2)

Based on Shannon’s theory, marginal entropies of the two random variables and their joint entropy are formulated in Equation 4, and 5 respectively.

\[
\begin{align*}
    H(B_1) &= \sum_{a} -P_{B_1}(a) \cdot \log P_{B_1}(a) \\
    H(B_1, B_2) &= \sum_{a,b} -P_{B_1,B_2}(a,b) \cdot \log P_{B_1,B_2}(a,b)
\end{align*}
\] (4) (5)

\( a \) and \( b \) are intensities of corresponding pixels in two images. \( P_{B_1}(.) \) describes the probability of a given intensity value being present in the image block ‘\( B \)’.

Phase Correlation (PC) is a frequency-domain based similarity measure that provides straightforward estimation of translation between two images, which is based on Fourier shift property (Yan and Li 2008). If \( B_1(i,j) \) and \( B_2(i,j) \) represent two image blocks and \( X_1[u,v] \) and \( X_2[u,v] \) show the corresponding discrete Fourier transforms (DFTs), the phase correlation is formulated as in Equation 6.

\[
S[n] = F^{-1} \left[ \frac{X_1[u,v] \times X_2^*[u,v]}{X_1[u,v] \times X_2^*[u,v]} \right]
\] (6)

where \( F^{-1} \) represents the inverse discrete Fourier transform. A window function e.g. Hamming is usually applied prior to Fourier transformation to remove edge effects and increase accuracy (Stone et al. 2001; Keller and Averbuch 2007).

If the two images are exactly the same, i.e., \( B_1(i,j) = B_2(i,j) \), phase correlation result is obtained as a peak of unity located at \( (i,j) = 0 \). However, in our case where the two images to be registered are not identical due to noise, intensity changes, etc., height of the peak of phase correlation is used as a good similarity measure for matching.

Phase correlation has another property that motivates its use in image registration. That is, if two images differ merely by translation, than the amount of this translation (or shift) can be obtained by phase correlation as formulated in Equation 7, and 8 respectively.

\[
B_2(i,j) = B_1(i-t_1, j-t_2) \\
S[i,j] = F^{-1} \left[ \frac{X_1[k] \times X_1[k] \times e^{-j\omega t} e^{-j\omega t'}}{X_1[k] \times X_1[k] \times e^{-j\omega t}} \right] = \delta[i-t_1, j-t_2]
\] (7) (8)

2.2. Evaluation measures for image registration

Inverse consistency: Consistency of registration can be evaluated by analyzing ‘invertibility’ property of transformation. ‘Forward transformation’ is described as transforming target to match reference, and ‘reverse transformation’ is described as transforming reference to match target image. Ideally ‘forward’ transformation should equal the ‘inverse of the reverse’ transformation producing a consistent correspondence between two images (Christensen et al. 2006). However, this is usually not the case for an actual image registration problem. Inconsistency occurs as a result of the matching criteria’s inability to uniquely describe the correspondence between two images (Christensen and Johnson 2001). Inverse consistency is quantified with an error statistic between a forward and reverse transformation, i.e. RMSE.

Global Similarity: Intensity based registration methods aim at optimizing image intensity similarities between the source and the target images. Global similarity assessment is based on the assumption that transformed target image is in good correspondence with the reference image. This correspondence can be quantified by obtaining similarity between images. High similarity is assumed to indicate good registration, although not fully guaranteed (Zitova and Flusser 2003).
3. Material and the study area

Hyperspectral spaceborne imaging in mission today is limited with NASA’s Earth Observing System EO-1 Hyperion sensor (URL-1) and ESA’s CHRIS/Proba (URL-2), an alternative research-oriented hyperspectral sensor that provides bands at VNIR range. EO-1 Hyperion has more than 200 spectral bands in VNIR and SWIR range with 30 m spatial resolution. There are numerous commercial spaceborne sensors that provide high spatial resolution images today. In this study, IKONOS multispectral image is utilized. IKONOS imagery (URL-3) has four multispectral bands with 4 m spatial resolution and a panchromatic band with 1 m spatial resolution.

In this study Hyperion bands in the visible range and IKONOS R, G, B bands with 4 m resolution is utilized. There is a spatial resolution ratio of 1/7.5 for high resolution IKONOS and low resolution Hyperion imagery. Hyperion bands in the visible range are used as high spectral resolution component of the registration process to obtain a deformation grid for geometric transformation that can be utilized to transform all of the remaining bands of Hyperion image in the NIR and SWIR ranges.

Images used in this study date back to 2007 summer and have 5 week temporal span of acquisition. Case area is near Durusu (Terkos Lake), 45 km North-West of Istanbul, and it covers a region of about 5 km$^2$ including agricultural areas, forests, Durusu settlement, and a part of the Black sea as shown in Figure 1.

Hyperion and IKONOS images being acquired from different sensors are subject to image multimodalities. These include: different spatial resolution, different spectral resolution, and intensity changes between the imageries due to time span in the growing season. IKONOS image is orthorectified. However, Hyperion Level 1gst image although reported as orthorectified (URL-1) has local distortions that are attributed to the low resolution Digital Elevation Model (DEM) used in orthorectification. Multimodalities of images from two different sensors that entail consideration in our case are given in Table 1.

4. Method

A pixelwise registration that is based on the similarity of reference and target blocks is proposed. IKONOS image is taken as target and Hyperion image is taken as reference and ground control for accuracy evaluation as well. Although IKONOS has true geometry with orthorectification, it is used as target and transformed to match Hyperion essentially not to alter radiometry of Hyperion that is going to be further used for spectral analysis. Translation at best match per Hyperion grid (30 m) was recorded and used as displacement vectors to construct a deformation grid. Deformation grid - similar to an optical flow - is used to establish a fine sampling of the deformation to be locally minimized. Subsequently, geometric transformation relocates each IKONOS pixel in a distance-weighted manner from four nearest displacement vectors in 30 m span and reassigns pixel values by interpolation.

Method is applied to multispectral R, G, B bands of IKONOS and synthetically produced R, G, B substitutes of Hyperion imagery. All of the processing has been implemented in MATLAB.

4.1 Preprocessing

4.1.1 Spectral Normalization

Hyperion imagery nominally has 10 nm bandwidth for each of its bands where IKONOS has broader multispectral response that corresponds to band 8-17, 16-27, and 27-36 of Hyperion R, G and B bands respectively. To make direct ra-
diometric comparison of both imageries, an aggregation that combines portions of the Hyperion bands to emulate Spectral Response Function (SRF) of IKONOS is needed. Gaussian curves that represent SRF of IKONOS sensor at R, G, B channels were used to produce weighted sum of Hyperion bands to produce synthetic Hyperion R, G, and B bands.

### 4.1.2 Spatial normalization

Hyperion and IKONOS imageries have to be brought into equivalent spatial resolution. IKONOS multispectral image with 4 m pixel size is downsampled to 30 m spatial resolution. Prior to downsampling, to overcome the aliasing effect in the frequency domain, a low-pass filter of window size that is twice the sampling rate is performed. For calculation of similarity measure for each translation, IKONOS image is downsampled to Hyperion resolution.

### 4.2 Translation of IKONOS imagery

#### 4.2.1 Determining block size

An adequate size for IKONOS blocks is defined for translation. Size of blocks should be large enough to produce meaningful statistics and not smaller than a particular size in order to capture image features. This critical threshold is determined considering spatial autocorrelation in the images based on the idea that adjacent observations of the same phenomenon are correlated (Griffith 2003). Extent to which there exists spatial autocorrelation is assumed to give a measure for feature sizes in the images. An empirical semivariogram that is based on common quantitative index of spatial autocorrelation; Geary Ratio (GR) (Geary 1954) and between-point distances is built. An exponential variogram model fitted to the empirical semivariogram is observed to level out at a particular interpoint distance that is called ‘range’ value. Range value represents distance where points (pixels) in the image are no longer autocorrelated and taken as empirical block size in this study. Range distance is approximately 950 meter that is equivalent to an area of about 31x31 pixels blocks of 30 m pixel size of Hyperion image.

#### 4.2.2 Translation

Blocks were translated 1 m each time within range -45 ±45 m. 45 m equals to one and a half pixel of Hyperion imagery, and it covers amount of misregistration between two images as we have observed. Translation is performed in both x, y directions from the origin for each reference pixel (90x90=8100 repetitions for each successive translation) (Figure 2). Downsampling phase in preprocessing step (section 4.1.2) is repeated after each translation. Three similarity measures; NCC, MI (including both MI and NMI), and PC were used to calculate similarity between target and the reference images for each translation. A windowing function is usually applied prior to PC to improve accuracy by reducing edge effects (Stone et al. 2001). For PC, a hamming window that is 2/3 of block size is utilized.

### 4.3 Construction of deformation grid

Translation index at x, y gives displacement direction and amount. Translation at best match is recorded per pixel as displacement vectors. This matrix of displacement vectors, so called optical flow, establishes a deformation grid that is used in geometric transformation (Figure 3). Deformation grid for this study has 30 m span in accordance with Hyperion image for its grids.

### 4.4 Geometric transformation

Last step is geometric transformation that utilizes deformation grid to transform original IKONOS imagery. Accordingly, displacement vectors that are regularly distributed at 30 m span were used to relocate each IKONOS pixel (4 m). For each IKONOS pixel, four nearest displacement vectors are distance weighted using Euclidian metric and a new displacement per pixel is set. This produces a fine deformation grid of 4 m sampling. Pixel values are reassigned according to this deformation grid by bilinear interpolation that produces smoother output compared to nearest neighbor interpolation.

---

Figure 2: Translation of target block on reference image for each reference pixel
Accuracy improvement and evaluation measures for registration of multisensor remote sensing imagery

Utilized to calculate similarity of the target IKONOS image have been utilized for registration; NCC, MI, and PC, this time are global similarity measures for transformation methods; NCC, NMI, PC. Therefore, MI results have been excluded. MI measures on the other side as opposed to reported results in the literature about higher performance in comparison to NCC for coping with intensity changes between images (Chen et al. 2003; Roshni and Revathy 2008) yielded very poor results in our study. PC yielded best results for all of the global similarity measures.

Process load is measured as the time elapsed for single run of a translation in a workstation and is given in Table 2. Accordingly; NCC has the least process load, where NMI has the highest process load that is about five times the NCC. PC has slightly more processing load compared to NCC.

### 6. Conclusions

In this study, we have proposed the use of ‘Inverse Consistency’ and ‘Global Similarity’ measures for evaluating accuracy of intensity based automated registration of Hyperion and IKONOS images using NCC, MI, and PC as similarity measures.

Experimental results although not giving an absolute accuracy measure, present accuracy measures for evaluating efficiency of the registration. Inverse Consistency provides an error measure about geometric correspondence, where Global Similarity gives a similarity measure of intensity correspondence. Global Similarity evaluations are based on the similarity of intensity values in transformed target and the reference. Therefore, difference in intensities among images is a factor that reduces the success of this evaluation to an extent method can cope with intensity changes. As a result, Global Similarity measure is prone to actual changes in intensity values. Invertibility property providing a geometric measure is more robust to these changes and represents method’s strength for accurate geometric transformation. Process load is another critical indicator for efficiency where larger images with many block translations may demand long periods of time of processing.

According to overall experimental results, PC show reasonably better evaluation results compared to the other methods. Process load for PC is slightly more than NCC, but this

### 5. Experimental Results

Three similarity measures that are commonplace in intensity-based image registration; NCC, MI (MI and NMI), and PC are examined for their performance and accuracy in our image registration problem. Accuracy and performance evaluations are made based on ‘inverse consistency’ of the transformation (Christensen et al. 2006) and ‘global similarity’ of the resultant images. In addition to these quantitative measures, performance is also tested in terms of process load that is quantified by time elapsed for the procedures in MATLAB.

**Inverse Consistency Results:** Inverse consistency for this study is reported as mean and root mean squared error (RMSE) between the forward and reverse transformations of the images (Table 2). According to error statistics obtained from inverse consistency assessment, PC with the lowest error score yielded most consistent transformation among registration results based on three similarity measures. NCC follows as the second best and NMI yielded unsatisfactory results with ME and RMSE far higher compared to other two methods.

**Global Similarity Results:** Similarity measures that have been utilized for registration; NCC, MI, and PC, this time are utilized to calculate similarity of the target IKONOS imagery and Hyperion imagery to assess accuracy of each registration. Evaluations were given as correlation coefficient ‘r’ for NCC, an entropy-based measure for NMI, and phase correlation coefficient for PC as depicted in Table 2. r for NCC can get values between 0-1 where 1 describes identical images. Entropy measure for NMI can get values between 0-2 where 2 describes identical images and PC coefficient can get values between 0-1 where 1 describes identical images. In the columns; NCC, MI and PC are expected to get higher values for good correspondence. Values are summed totals of similarity measures for three bands; R, G and B.

According to measures obtained from global similarity assessment, MI yielded poor registration accuracy with lowest similarity between transformed target and the reference image. MI has shown results of poorer quality compared to NMI as previously reported (Pluim et al. 2003; Suganya et al. 2010). Therefore MI results have been excluded. MI measures on the other side as opposed to reported results in the literature about higher performance in comparison to NCC for coping with intensity changes between images (Chen et al. 2003; Roshni and Revathy 2008) yielded very poor results in our study. PC yielded best results for all of the global similarity measures.

### Table 2: Inverse consistency error statistics and global similarity measures for transformation methods; NCC, NMI, PC.

<table>
<thead>
<tr>
<th>Similarity Measures</th>
<th>Forward-Backward Transformation Difference</th>
<th>Global Similarity</th>
<th>Process Load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Mean Error</td>
<td>NCC</td>
</tr>
<tr>
<td>NCC</td>
<td>6.9590</td>
<td>3.2101</td>
<td>2.7218</td>
</tr>
<tr>
<td>NMI</td>
<td>9.9147</td>
<td>3.9068</td>
<td>2.5056</td>
</tr>
<tr>
<td>PC (2/3 windowing)</td>
<td>*1.8002</td>
<td>*1.6721</td>
<td>*2.7908</td>
</tr>
</tbody>
</table>

(* Lowest error measure for inverse consistency, highest similarity measure for global matching, and shortest time for processing)
is tolerable regarding the improvement it has made in the results. NCC has slightly worse evaluation results compared to PC. However, NMI yielded quite unsatisfactory results where MI yielded results even worse. Overall scores obtained for MI and NMI in our study in terms of accuracy and process load discourages the use of these methods in intensity-based registration of multisensor remote sensing imagery.

Evaluations provide useful information to decide for the method to employ. However it should be noted that evaluations on accuracy of image registration is a nontrivial task. Inaccuracy might be due to errors that get involved into image registration at some processing stage. Moreover, actual physical differences in the image contents might also be attributed as inaccuracy.

Acknowledgements

This study has been supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) as part of Project: 111E052.

References


Griffith D.A., (2003), Spatial Autocorrelation and Spatial Filtering, Springer Berlin, Germany.


Roshni V.S., Revathy Dr. K., (2008), Using mutual information and cross correlation as metrics for registration of images, Journal of Theoretical and Applied Information Technology, 4(6), 474-481.


