MULTITEMPORAL IMAGE REGISTRATION BASED ON
MULTIRESOLUTION DECOMPOSITION

Registro de Imagens Multitemporais Baseadas em Decomposição Multiresolução

Leila M. G. Fonseca ¹
Max H. M. Costa ²
Thales S. Korting ¹
Emiliano Castejon ¹
Felipe C. Silva ¹

¹ National Institute for Space Research – INPE
Image Processing Division
C.P. 515, 12227-001, São José dos Campos (SP), Brazil
{leila, tkorting, felipe, castejon}@dpi.inpe.br

² State University of Campinas - Unicamp
FEEC/DECOM
C.P. 6101, 13081-970 Campinas (SP), Brazil
max@decom.fee.unicamp.br

ABSTRACT

Image registration is the process of overlaying two or more images of the same scene taken at different times, from
different viewpoints, and/or different sensors. With the increase in the number of images collected every day from
different sensors, automated registration of multi-sensor, multi-spectral and multi-temporal images has become an
important application. In this paper an efficient algorithm for the registration of multi-temporal images with similar
spectral responses is proposed. The procedure relies on the gray level information content of the images and their local
wavelet transform modulus maxima. The registration algorithm is performed at progressively higher resolutions, which
allows for faster implementation and higher registration precision. For the matching process, a novel procedure for
automatic point feature detection based on wavelet transform is presented. Moreover, an investigation about the
performance of four similarity measures in the task of feature point matching for satellite images is also presented. The
results show that the matching method based on correlation is efficient to register images of similar spectral responses
even if multi-temporal changes are present. To overcome the presence of rotation angle between the images to be
registered, a procedure to estimate this angle is performed before the matching process. Measurements of performance
are provided, as well as extensive experimental results, to illustrate the quantitative performance of the registration
algorithm. Images of different sensors and acquisition time are used in the experiments.

Keywords: Image Registration, Wavelet Transform, Matching, Feature Detection

1. INTRODUCTION

Agencies from all over the world have launched many earth observation systems of high
resolution, multi-bands and multi-sensors. The analysis
of multitemporal and multisensor remote sensing
datasets acquired by these systems can be efficiently
done if the data refer to a common geometry. The
processing of images such that their corresponding
pixels have the same geometry comprises registration or
geocoding of the images.

The high resolution of the sensors means that
images must be done with considerable precision. The
increased volume of data motivates development of
automatic or quasi-automatic registration systems.
Development of automatic image registration methods
is still a wide field of research, due to the need to use
them with images which vary significantly in content,
 radiometry and geometry. The general approach to
image registration consists of the following four steps
(Fonseca and Manjunath, 1997):

1. Feature identification: identifies a set of
 relevant features in a pair of images, such as
 edges, line intersections, region contours,
 regions textures, etc.
2. Feature matching: establishes correspondence between the identified features. Thus, each feature in the sensed image must be matched to its corresponding feature in the reference image. Each feature is identified with a pixel location in the image. The corresponding points are usually referred to as control points.

3. Spatial transformation: determines the mapping function that can match all the points in the image using information about the control points obtained in the previous step.

4. Interpolation: resamples the sensed image using the mapping functions to bring it into alignment with the reference image.

In general, registration methods differ from each other in that they can combine different techniques for feature identification, feature matching, and mapping and interpolation functions. The most difficult step in image registration is obtaining the correspondence between the two sets of features. This task is crucial to the accuracy of image registration and much effort has been spent in the development of efficient feature matching techniques. Given the matches, the task of computing the appropriate mapping functions does not involve much difficulty. The interpolation process is also quite standard.

Generally, there are two classes of methods that can be used to register images: area based and feature based matching (Fonseca and Manjunath, 1996; Zitova and Flusser, 2003). Area based methods can be applied to spectrally similar images whilst feature based matching can be used more generally to register any pair of images. Several image registration methods have been proposed in the literature (Li et al. (1995), Govindu et al. (1998), Fonseca et al. (1998), Dare and Dowman, (2001), Kanade and Okutomi, (1994), Cole-Rhodes et al., (2003)). Fookes et al. (2004) proposed extensions to Mutual Information (MI) based stereo matching in order to increase the robustness of the MI algorithm. Fedorov et al., (2002) developed an operational system for automatic image registration, using three approaches found on literature (Kenney and Majunath, (2001), Fonseca et al., (1998) and Li et al., (1995)), including also techniques for mosaicking images.

In this paper we present an efficient approach for registration of multi-temporal images with similar spectral responses. The correlation coefficient is used as a similarity measure and only the best pairwise fittings among all pairs of feature points are taken as control points. A consistency checking step is also involved to eliminate mismatches. This way we have a reliable initial guess for the registration transformation which is a crucial phase in the process. The wavelet transform decomposition is used to extract feature points which are taken as control points and to decompose the images in different resolution levels. The algorithm is performed at progressively higher resolution, which allows for faster implementation and higher registration precision.

The registration algorithm is very simple and easy to apply because it needs basically only one parameter. Because the matching is carried out only on the selected feature points and in a coarse-to-fine manner, a significant amount of computation is saved in comparison to traditional pixel-by-pixel searching methods. Due to the fact that the registration procedure uses the gray level information content of the images in the matching process it is more adequate to register images of the same sensor or with similar spectral bands. In spite of this, it has demonstrated technical feasibility for many images of forest, urban and agricultural areas from Thematic Mapper (TM-5), SPOT and radar sensors taken at different times.

In addition to the registration method presented, one of the main contributions of this paper is the procedure for point feature detection based on wavelet decomposition. The method automatically selects point features to be used in the matching process. It is very simple and reasonably efficient in terms of computational complexity. Moreover, an investigation about the performance of four similarity measures in the task of feature point matching for satellite images is also presented. The results show that the matching algorithm based on correlation is efficient to register images of similar spectral responses even when multi-temporal changes occur. In order to overcome the presence of rotation angle between the images to be registered, a procedure to estimate this angle is performed before the matching process.

The organization of the paper is as follows. In the next section we provide the reader with an introduction to wavelet transforms. In the subsequent sections, we discuss in detail each step of the registration algorithm. Following this we present some results of registering satellite images. Finally, we present some conclusions.

2. WAVELET TRANSFORMS

Although the Wavelet theory is well known in the literature, we will briefly describe it in order to facilitate the description of equations in the posterior sections. Interested readers can find more information on wavelets and the wavelet transform in Chui, (1992), Fliege (1994), Graps, (1995) and Mallat (1989a, 1989b, 1999).

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions.

Let \( \Psi \in L^2(\mathbb{R}^n) \) be a complex valued function. If the function \( \Psi \) satisfies

\[
\int_{\mathbb{R}^n} \Psi(x) dx = 0
\]  

(1)
it may be called a basic wavelet or “mother wavelet”. The families of functions
\[ \psi_{a,b}(x) = a^{-\frac{1}{2}} \psi \left( \frac{x-b}{a} \right) \]
with \( a, b \in \mathbb{R}, a \neq 0 \) \hspace{1cm} (2)
generated from the “mother wavelet” under the operations of dilation (or scaling) by a factor \( a \) and translation in time by a parameter \( b \) form a wavelet family.

The continuous wavelet transform of a function \( f x \in L^2 \mathbb{R}^2 \) with respect to this family of wavelets is given by the convolution integral
\[ W_a f x = \int f u \psi \left( \frac{u-x}{a} \right) du \]
where \( \psi_a x = a^{-\frac{1}{2}} \psi \left( \frac{x}{a} \right) \).

The wavelet can be interpreted as the impulsive response of a band-pass filter, and the wavelet transform of a function as a convolution of the function with the dilated filter responses (GRAPS, 1995).

In practice the scale \( a \) should be discretized. For a particular class of wavelets, the scale \( a \) can be sampled along a dyadic sequence \( a = 2^j \) with \( j \in \mathbb{Z} \), without modifying the overall properties of the transform (Mallat and Zhong, 1992). The transform corresponding to dyadic values of \( a \) is called the discrete wavelet transform (DWT), given by

\[ W_a f x = \int f u \psi \left( \frac{u-x}{a} \right) du \]
where \( \psi_a x = a^{-\frac{1}{2}} \psi \left( \frac{x}{a} \right) \).

The wavelet transform of 1-D functions can be extended to the 2-D case. Let us consider a smoothing function \( \psi x, y \), taken as the impulse response of a 2-D low-pass filter. The first order derivative of \( \psi x, y \), decomposed in two components along the \( x \) and \( y \) directions, respectively, are
\[ \psi^1 x, y = \frac{\partial \psi}{\partial x} \]
\[ \psi^2 x, y = \frac{\partial \psi}{\partial y} \]

and these functions can be used as wavelets.

For any 2-D function \( f A \times A \), the wavelet transform at scale \( a = 2^j \) defined with respect to these two wavelets has two components (Mallat and Hwang, 1992), given by

\[ W_{a^j} f x, y = f x, y \frac{\partial \psi}{\partial x} \]
\[ W_{a^j} f x, y = f x, y \frac{\partial \psi}{\partial y} \]

Therefore, these two components of the wavelet transform are proportional to the coordinates of the gradient vector of \( f x, y \) smoothed by \( \frac{\partial \psi}{\partial x} \) and \( \frac{\partial \psi}{\partial y} \). They characterize the singularities along \( x \) and \( y \) directions, respectively (Mallat and Hwang, 1992).

We define the function

\[ D b M f 2^j x, y = \int W_{a^j} f x, y \]

which is the modulus of the wavelet transform at the scale \( 2^j \). One can prove that for wavelets defined by Equation 6, \( D_{a^j} f 2^j x, y \) is proportional to the magnitude of the gradient field (Mallat and Zhong, 1992).

The angle \( a \) between the gradient vector of \( f x, y \) and the horizontal is given by

\[ a = \tan^{-1} \left( \frac{W_{a^j} f x, y}{W_{a^j} f x, y} \right) \]

This angle indicates locally the direction where the signal exhibits the sharpest variations.

3. FEATURE POINT DETECTION USING WAVELET TRANSFORM

This section presents a method to select automatically point features to be used in the matching process. The features are extracted with the use of the modulus maxima of the wavelet transform. Moigne, (1994), Djamdji, (1995) and Corvi and Nicchiotti, (1995) also use the wavelet transform to extract salient features in the images, which can be used in the registration process. Moigne, (1994) uses a percentage.
(10%) of the total number in bands HL and LH of the wavelet decompositions to represent the significant features in the images. Djamdji, (1993) use the points in which wavelet coefficients à trous (BIJAOUI and GIUDICELLI, 1991) are maximum. Corvi and Nicchiotti, (1995) use the points in which wavelet coefficients are maximum or minimum values. The matching processing is performed separately for maxima and minima features. Zheng and Chellappa, (1993) use Gabor decomposition for feature extraction, which is computationally more expensive and complex.

Differently, we use the local maxima of wavelet transform to extract point features in the image. Our implementation uses three filters G, L and H, derived from cubic spline function, whose coefficients are shown in Table 1. The two components of the wavelet transform in the x and y directions are used to calculate the gradient field. The edge information present in the high frequency subbands of the wavelet decomposition we use an algorithm proposed by Mallat and Zhong, (1992), which is rather efficient in terms of computation time. Memory reduction is an important factor in remote sensing image processing since the amount of data is very large.

### TABLE 1- FILTERS CORRESPONDENT TO CUBIC SPLINE WAVELET

<table>
<thead>
<tr>
<th>n</th>
<th>G</th>
<th>L</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.00008</td>
<td>0.00003</td>
<td>0.0625</td>
</tr>
<tr>
<td>1</td>
<td>-0.01643</td>
<td>0.00727</td>
<td>0.2500</td>
</tr>
<tr>
<td>2</td>
<td>-0.10872</td>
<td>0.03118</td>
<td>0.3750</td>
</tr>
<tr>
<td>3</td>
<td>-0.59261</td>
<td>0.06623</td>
<td>0.2500</td>
</tr>
<tr>
<td>4</td>
<td>0.59261</td>
<td>0.79113</td>
<td>0.0625</td>
</tr>
<tr>
<td>5</td>
<td>0.10872</td>
<td>0.06623</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.01643</td>
<td>0.03118</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.00008</td>
<td>0.00727</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.00003</td>
<td></td>
</tr>
</tbody>
</table>

Let us call the image to be warped the warp image and the image to which the warp image will be reduced the reference image. We compute the discrete multiresolution wavelet transform (L levels) of the two images.

The wavelet decomposition of an image is similar to a quadrature mirror filter decomposition with the low-pass filter L and its mirror high-pass filter H (FLIEGE, 1994). We call LL, LH, HL and HH the four images created at each level of decomposition, as in (MOIGNÉ, 1994), where this decomposition is also used for the purpose of image registration. In the current implementation, filters G and L are used to calculate bands LH and HL, and filter H is used to calculate band LL. Band HH has not been used in our implementation.

The next phase aims to identify features that are present in both images in each level of the decomposition. Here we use the modulus maxima of the wavelet transform to detect sharp variation points which correspond to edge points in the images. The LH and HL subbands at each level of the wavelet transform are used to estimate the image gradient. For the wavelet decomposition we use the filters given in Cheong et al., (1992).

In the feature point selection process, in each level of decomposition, four steps are involved:

- Edge points extraction using the modulus of the wavelet transform;
- Selection of edge points localized in high contrast regions;
- Suppression of non-maximum local edge points.

In the first step, the reference and warp images are processed in L levels in the wavelet decomposition. In each resolution level, the modulus of the wavelet transform $M_{j}$ and the angle $\alpha_{j}$ are computed.

A point $P(x^{\alpha}, y^{\alpha})$ in resolution level $j$ is recognized as an edge if

$$M_{j}(x^{\alpha}, y^{\alpha}) > M_{j-1}(x^{\alpha}, y^{\alpha})$$

where $x^{\alpha}, y^{\alpha}$ belongs to the two nearest neighbors of $P(x^{\alpha}, y^{\alpha})$ in the direction to which the gradient vector $\alpha_{j}$ points.

A thresholding procedure is applied on the wavelet transform modulus image in order to eliminate non-significant edge points. Thus, a point $x^{\alpha}, y^{\alpha}$ is recorded only if

$$M_{j}(x^{\alpha}, y^{\alpha}) > \tau_{j}$$

where $\tau_{j}=\beta \sigma_{j} + \mu_{j}$ is a constant whose initial value is defined by the user and $\sigma_{j}$ and $\mu_{j}$ are the standard deviation and mean of the wavelet transform modulus image at level $2^{j}$, respectively. The parameter $\beta$ controls the number of feature points selected for the matching. Since the number of feature points increases at finer resolutions, the parameter $\beta$ is also increased in the higher levels of decomposition in order to select the most significant feature points in the images. This condition eliminates weak edges and ensures a consistency of edges in all resolution levels. Therefore, if the edge is sufficiently strong, it will appear practically in all resolution levels. This parameter is defined by user.

Consider a window $w$, $B w$ ($w$ is an odd integer) of points centered in $P(x_{o}, y_{o})$. Let $\sigma_{w}$ be the standard deviation of the image values inside the window. Let $CO(x^{\alpha}, y^{\alpha})$ denote a contrast measure defined as (Gonzalez and Woods, 1992):

$$CO(x_{o}, y_{o}) = \frac{1}{1 + \sigma_{w}}$$

Revista Brasileira de Cartografia No 60/03, outubro 2008. (ISSN 1808-0936) 274
In order to ensure that feature points are localized in high contrast regions, a point \( P \ x_0, y_0 \) is recorded if and only if

\[
CO \ x_0, y_0 \geq T_w
\]  

(12)

where \( T_w \) is a given threshold above which the contrast measure is considered sufficiently large. This condition reduces the number of mismatched pairs in the matching process, which is discussed in the next section. Finally, the local maxima are calculated as

\[
M_{ij} * x_0, y_0 = \max_{x,y \in V_p} M_{ij} * x, y
\]  

(13)

where \( V_p \) is a neighborhood of \( P \ x_0, y_0 \), defined as a window of \( w_x \) by \( w_y \) points centered at \( P \ x_0, y_0 \). This condition ensures that the feature point is unique in its surrounding area.

4. POINT MATCHING

In this section, we shall verify the correct matching of the selected feature point pairs in the reference and warped images. A correlation based matching measure is used in the matching process. The use of the correlation coefficient as a similarity measure is motivated by good results obtained in an experimental study. This study is briefly presented in the next section.

4.1 Defining a similarity measure

Area based matching methods use similarity measures to compare windows of two images. Various similarity measures have been reported in the literature (Pratt, 1974), Barnea and Silverman, (1972), Khosravi and Schafer, (1996) and Brunelli and Messelodi, (1995). Brunelli and Messelodi, (1995) have compared the correlation coefficient to similarity measures based on the \( L_1 \) norm. In tasks involving face recognition, similarity measures based on the \( L_1 \) norm outperformed the correlation measure. In order to verify the performance of some similarity measures in the task of feature points matching we have developed a comparative study with four similarity measures. These include the traditional correlation coefficient, the Gray Hit or Miss Transform (GHMT) (Khosravi and Schafer, 1996), the Sequential Similarity Detection (SSD), and one of the measures based on the \( L_1 \) norm (L1G) proposed by Brunelli and Messelodi, (1995).

Let \( f \) and \( g \) be two grayscale images. Let \( t \) be a \( LBK \) window of \( g \) centered at \( m, n \), called a \( b \) by \( c \) template. Let \( W = i, j : i = 0, ..., L \ @ 1, j = 0, ..., K \ @ 1 \) be the support of \( t \). The comparison between the template and a window of \( f \), with support \( W \) and centered at \( x, y \), is accomplished through similarity measures. The following equations define the four similarity measures mentioned above.

**Correlation Coefficient:**

\[
C \ x, y = \frac{\sum_{i,j} (f_{x+i,y+j} - \mu_f)(g_{x+i,y+j} - \mu_g)}{\sqrt{\sum_{i,j} (f_{x+i,y+j} - \mu_f)^2 \sum_{i,j} (g_{x+i,y+j} - \mu_g)^2}}
\]

(14)

where \( \mu_f \) and \( \mu_g \) are the local means (average intensity values) of \( f \) and \( t \), respectively. \( W \) is the support of template \( t \), and \( LBK \) is the cardinality of \( W \).

The Gray Hit or Miss Transform (GHMT):

\[
M \ x, y = \min_{i,j \in W} \left( f_{x+i,y+j} - \mu_f \right) \left( g_{x+i,y+j} - \mu_g \right)
\]

(15)

**Sequential Similarity Detection (SSD):**

\[
S \ x, y = \max_{i,j \in W} \left( f_{x+i,y+j} - \mu_f \right) \left( g_{x+i,y+j} - \mu_g \right)
\]

(16)

\[
L_1 \ Norm (L1G):
\]

\[
G \ x, y = \left( f_{x+i,y+j} - \mu_f \right) \left( g_{x+i,y+j} - \mu_g \right)
\]

(17)

where \( f \) and \( t \) are normalized values and \( G \ x, y \geq 0, 1 \).

We have investigated the performance of these measures in the task of feature point matching for actual satellite images. In the experiment, we selected 6 pairs of multitemporal images of size 512x512 pixels: two of urban, two of forest and two of agricultural areas. These images are selected for the diversity of information that they presented, allowing for a more robust validation. The images were decomposed in two resolution levels and the matching process was accomplished in the lowest resolution level of the LL band. This procedure was adopted because it simulates the way in which the proposed registration algorithm is implemented.

One way to validate a measure of similarity is by quantifying its ability to detect correct matches. The larger the number of correct matches the better the performance of the corresponding similarity measure. When we use the correlation coefficient, for instance, to compare two features, the maximization of this measure for selecting the best match does not assure that the match obtained is correct. Although the correlation value is maximum, their low absolute values may lead

Revista Brasileira de Cartografia No 60/03, outubro 2008. (ISSN 1808-0936) 275
to false matched points. In this case, we say the matching has low confidence. Problems such as noise, occlusion and low contrast in the images can lead to mismatches. The absolute value of the correlation coefficient itself may be used as a meaningful confidence measure. Thus, a pair of feature points will be matched if the absolute value of the correlation coefficient is larger than a certain threshold. To validate the measures of similarity, we quantify their capacity to detect correct matches. A good similarity measure is one which gives very little or no false alarms and a minimum number of misses in the matching process. A false alarm is defined as a false match declared as valid, whereas a miss is defined as a valid match declared as false.

The experimental investigation has shown that the correlation coefficient outperformed the other methods for the chosen set of parameters. Fig. 1 plots the percentage of correct matches against the size $M$ of the matching window for all methods. One can observe that GHMT method presents some oscillations with a salient peak around $M=13$. This behavior indicates a certain sensitivity of this method with respect to the window size. The correlation and $L_1G$ methods present similar performances and outperform the other methods, mainly for window sizes 6 or greater.

It must also be considered that the $L_1G$ method is computationally more intense than the correlation method. Based on these evidences, the correlation coefficient measure demonstrated good robustness and efficiency. Therefore, it was chosen for use in the proposed registration method.

It is known that area based matching methods are very sensitive to rotation. Zheng and Chellappa, (1993) and Hsieh, (1997) have proposed different approaches to overcome this problem. Zheng and Chellappa, (1993) use a shape-from-shading technique to estimate the illuminant directions of images under the assumption that the illumination source is stationary. By taking the difference between the illuminant directions, the rotation angle between images can be estimated. Hsieh, (1997) uses a line-fitting model to estimate the directions of all the edges from the wavelet transform maxima. In order to estimate the orientation differences between the images, a so-called “angle histogram” is calculated. Then, the rotation is found by seeking the angle that corresponds to the maximum peak in the histogram. Both methods have drawbacks. As pointed out by Hsieh (1997), Zheng and Chellappa’s approach does not work satisfactorily if the scene includes many buildings and objects due the fact that the illumination conditions in one image may not be equivalent to those in the other. On the other hand, Hsieh’s method can fail if the changes between the images are reasonable high.

Fig. 1 – Performance of four matching methods with respect to matching window size: correlation coefficient (dashed line), GHMT, SSD, and $L_1G$.

Fig. 2 shows the percentage of false alarms against the number of correct matches with window size $M=13$ for the correlation and $L_1G$ methods. One can observe that correlation still slightly outperforms $L_1G$.

Fig. 2 – False alarm rate versus number of correct matches for correlation coefficient and $L_1G$ methods (with $M=13$).

The method proposed in this paper aims mainly at registering remote sensing images, and the angle of rotation between satellites can be readily obtained as auxiliary information. If it is not the case, the proposed
scheme estimates the rotation angle by choosing windows about each feature point that have been rotated so that their central gradient points downward (Kenney and Manjunath, 2001).

4.2 Initial Point Matching

The feature matching process is performed through a combination of area and feature-based techniques. Fig. 3 shows the block diagram and the steps involved in the initial matching process. The feature point matching is achieved by maximizing the correlation coefficient over small windows surrounding the feature points within the LL subbands of the wavelet transform. Let the LL subbands of the warp and reference images be \( f_x \) and \( f_y \), respectively, and let \( C_{f_x, f_y} \) be the corresponding correlation coefficient.

\[
C_{f_x, f_y}(x, y, X, Y) = \max_{1 \leq i \leq N_x, 1 \leq j \leq N_y} C_{f_x, f_y}(x_i, y_i, X_j, Y_j) \tag{18}
\]

Therefore, the matching process is achieved in the following way. For each point \( f_x, x_i, y_i, 2 P_x \), all points \( f_y, X_j, Y_j, 2 P_y \) are examined and the most similar point \( f_y, X_k, Y_k \) is chosen. Next we test whether the achieved correlation is reasonably high.

If \( C_{f_x, f_y}(x_i, y_i, X_k, Y_k) > T_r \), then \( f_y, X_k, Y_k \) is called “the best match” of \( f_x, x_i, y_i \).

To verify that the match is consistent in the reverse direction, we test whether the best match of \( f_x, X_k, Y_k \) exists and is indeed given by \( f_y, x_i, y_i \). If that is the case, both points are taken as matched.

This reverse verification reduces the number of mismatched pairs in the matching process and allows for the use of smaller window sizes. Nevertheless, some false matches will inevitably occur. This initial part of the matching process is a crucial phase of the registration process. If the initial registration parameters are invalid the search for a registration transformation evolves in a wrong direction, and the correct trend may not be recovered in later steps. Therefore, a consistency-checking procedure is performed in order to eliminate incorrect matches and to improve registration precision.

Consistency-checking approaches were considered for use in the proposed algorithm. The first approach was similar to the one used in Li et al. (1995), which is performed recursively in such a way that the most likely incorrect match is deleted first, followed by the next most likely incorrect match, and so on. This approach is based on the fact that distances are preserved under a rigid transformation. In order to handle the situation in which the deformation between the images is not rigid, we propose an alternative procedure. This is an empiric method that uses combinatorial search. Let be two pairs of control points \( P_k = X_k, Y_k, x_k, y_k \) and \( P_l = X_l, Y_l, x_l, y_l \).

Let be \( D_{ik} \) and \( d_{il} \) the Euclidean distances between the points \( P_k \) and \( P_l \) in the reference and adjust images, respectively:

\[
D_{ik} = \min \|X_i \otimes X_k + Y_i \otimes Y_k\|_2 \tag{19}
\]

\[
d_{il} = \min \|X_i \otimes x_k + Y_i \otimes y_k\|_2 \tag{20}
\]

Let \( E_{ij} \) a error measure that relates the Euclidean distances \( D_{ij} \) and \( d_{ij} \) between the control points \( P_i \) and \( P_j \):
for $i,j=1,...,K$ and $i \neq j$. Let $p$ the minimum number of control points necessary to calculate the parameters of the transformation function selected and $T_x$ an error threshold. The outliers elimination procedure as follows:

(i) Group the $K$ control points pairs in the list of candidate pairs $L_K$ in sets of $n=p+1$ elements, denoted by $C_i^n = P_1, P_2, ..., P_n$, $i = 1, ..., n! \ K \frac{1}{n!}$;

(ii) For each set $C_i^n$, calculate the error, $E_i^n$, using the control point pairs in the set;

(iii) Select the set $C_i^n$ with the least error $E_i^n$;

(iv) For every $P_j \in L_K$ and $P_j \in C_i^n$, $j = 1, ..., K$ do:
   - Add a pair of control points $P_j$ to the set $C_i^n \cup C_i^n$;
   - Calculate the set error for $C_i^n$, denoted by $E_i^n$;

(v) Select the set $C_i^n$, among all sets generated in the previous step, with the least error $E_i^n$;

(vi) if $E_i^n > T_x$, stop;

(vii) Let $C_i^n = C_i^n$;

(viii) Go to step iv;

The possibility that the $p+1$ elements from initial set are wrong and even so present a small initial error exists. But, in this case, as points are added, the error tends to increase rapidly and, thus, one can detect the unfavorable situation interrupting the process and restarting it by choosing the next set with least error. In all tested cases in this work, this method was successful to perform consistency checking.

4.3 Image Warping

The above procedure provides a set of reliable matches which are used to determine a warping function that yields the best registration of the LL subbands at the precision available in level $L$ of the wavelet transform.

To model the deformation between the images, a 2D affine transform with the parameters $s, \theta, \Delta x, \Delta y$ is used (ZHENG and CHELLAPPA, 1993), such that

$$
X = T_1 x, y = s x \cos \theta + y \sin \theta + \Delta x
$$

$$
Y = T_2 x, y = s \cos \theta + y \sin \theta + \Delta y
$$

(21)

where $x, y$ and $X, Y$ are corresponding points in the warp and reference images, respectively. This model is commonly used in remote sensing applications and is a good approximation for images taken under similar imaging directions (Dana and Anandan, 1995), and which have been geometrically corrected (e.g. for Earth curvature and rotation). In most remote sensing applications the images have a similar level of geometrical correction, which enables the use of this class of transformations.

4.4 Refinement

The point matching and image warping steps can be performed at progressively higher resolutions in a similar fashion to that described above. At each level $l < L$, the warp image $f_r$ is transformed using the parameters estimated from the wavelet transform at the lower resolution level $(l-1)$. In other words, the LL, LH, HL, and HH subbands at level $l+1$ are recombined to reconstruct the LL subband at level $l$, which is then warped by the transformation specified at the previous point matching operation.

Fig. 4 shows the block diagram of the steps involved in the registration refinement process. Let $f_r$ denote the registered image. Registration refinement is achieved using the registered image and the set of feature points detected in the reference image. Each feature point $f_r, X_k, Y_k$ detected in the image $f_r$ at level $l$ is matched to $f_s, x_l , y_l$ if

$$
C_{f_r, f_s} = \max_{2 \leq m, n \leq 2} C_{f_r, f_s} X_k, Y_k = X_l + m, y_l + n, X_k, Y_k
$$

(22)

where $w_r$ is the width of the registration refinement window.

The traditional measure of registration accuracy is the root mean square error ($RMSE$) between the matched points after the transformation, defined as

```
Revista Brasileira de Cartografia N° 60/03, outubro 2008. (ISSN 1808-0936) 278
```
where $N$ is the number of matched points. This measure is used as a criterion to eliminate earlier matches which are considered imprecise. Poor matches are sequentially eliminated in an iterative fashion until the $RMSE$ value is lower than 0.5 pixel.

In the refinement process, the matching operation is performed only for those points, which fall on the overlapping region of the reference and warp images. This eliminates unnecessary computations and speeds up the procedure.

At each level the warping parameters are updated considering the parameters obtained in the previous level and in the refinement step. After processing all levels the final parameters are determined and used to warp the original warp image, thereby producing the final registration of the warp image with respect to the reference image.

5. EXPERIMENTAL RESULTS

The feature extraction algorithm was tested for many types of images. Next, three experiments are presented. The images used in the experiments correspond to a forest region in Amazon (band 5), an urban area in the Brasilia city (band 4), and an agriculture region close to Agudos city (band 5), acquired by the satellite Landsat-TM5. The original images (512x512 pixels) were processed for two decomposition levels.

Figures 5, 6 and 7 show the distribution of point features superimposed on the band LL of the wavelet decomposition in the two resolution levels and on the original test images. The parameters used in all experiments were: $\beta = 3$ (to eliminate less significant features), $T_w = 0.95$ for the contrast threshold and a neighborhood $V_p$ of 7x7 pixels in the local maxima selection process. These parameters values have been successfully tested for various images, which indicates the reliability of the method.

From this point on, we present some registration results of simulated images in order to show the performance of the registration algorithm proposed in this work. First, the images are artificially deformed using known transformation parameters (rotation, translation and scale). From this set of parameters we can calculate the actual required transformation to register the images. Next, the deformed images are transformed back (registered) using the proposed registration algorithm. This transformation process provides estimated parameters, which give very good approximations to the actual transformations used in the experiment, as seen in Table 2.
Fig. 6 – Selected feature points of an urban area (Brasilia), at resolution (a) 128x128, (b) 256x256, and (c) 512x512 pixels.

Fig. 7 – Selected feature points of an agriculture area (Agudos), at resolution (a) 128x128, (b) 256x256, and (c) 512x512 pixels.
For this experiment we selected 6 images of different regions and sensors. Figures 8(a)-(f) show the original images and Table 2 shows the distortion model parameters (actual and estimated transformations).

**TABLE 2 – ACTUAL AND ESTIMATED DISTORTION PARAMETERS**

<table>
<thead>
<tr>
<th>Images</th>
<th>Area</th>
<th>Parameters</th>
<th>s</th>
<th>θ</th>
<th>Δx</th>
<th>Δy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES-1 (radar)</td>
<td>agric/forest</td>
<td>actual</td>
<td>1.1</td>
<td>9.0</td>
<td>100.0</td>
<td>70.0</td>
</tr>
<tr>
<td></td>
<td>600×600</td>
<td>estimated</td>
<td>1.1</td>
<td>8.99</td>
<td>99.68</td>
<td>69.97</td>
</tr>
<tr>
<td>aerial photograph</td>
<td>urban</td>
<td>actual</td>
<td>1.0</td>
<td>-12.0</td>
<td>-100.0</td>
<td>-50.0</td>
</tr>
<tr>
<td></td>
<td>600×600</td>
<td>estimated</td>
<td>1.0</td>
<td>-12.0</td>
<td>-100.1</td>
<td>-49.92</td>
</tr>
<tr>
<td>SIR-C (radar)</td>
<td>agricultural</td>
<td>actual</td>
<td>0.92</td>
<td>8.0</td>
<td>80.0</td>
<td>-20.0</td>
</tr>
<tr>
<td></td>
<td>512×512</td>
<td>estimated</td>
<td>0.919</td>
<td>8.0</td>
<td>79.97</td>
<td>-19.92</td>
</tr>
<tr>
<td>optical sens (balloon)</td>
<td>desert</td>
<td>actual</td>
<td>0.90</td>
<td>15.0</td>
<td>38.0</td>
<td>-55.0</td>
</tr>
<tr>
<td></td>
<td>512×512</td>
<td>estimated</td>
<td>0.90</td>
<td>14.99</td>
<td>37.98</td>
<td>-54.97</td>
</tr>
<tr>
<td>Landsat TM</td>
<td>forest</td>
<td>actual</td>
<td>1.10</td>
<td>20.0</td>
<td>45.0</td>
<td>-100.0</td>
</tr>
<tr>
<td></td>
<td>1000×1000</td>
<td>estimated</td>
<td>1.10</td>
<td>20.01</td>
<td>45.12</td>
<td>-100.1</td>
</tr>
<tr>
<td>SPOT</td>
<td>vegetation</td>
<td>actual</td>
<td>0.95</td>
<td>10.30</td>
<td>385.0</td>
<td>201.0</td>
</tr>
<tr>
<td></td>
<td>1000×1000</td>
<td>estimated</td>
<td>0.95</td>
<td>10.29</td>
<td>384.56</td>
<td>200.63</td>
</tr>
</tbody>
</table>

Fig. 8(a) is a radar image AES-1 (Aerosensing Radarsystems, Germany) of an agriculture and forest area in Wessling region, band X, 600B 600 pixels, resampled for 2 meter pixel size (original resolution = 0.5 m). Shown in Fig. 8(b) is an aerial photograph of the city of São José dos Campos, SP, Brazil, 600B 600 pixels, taken on 04/08/1976. A radar image SIR-C/X-SAR (Spaceborne Imaging Radar C/X-Band Synthetic Aperture Radar) acquired on 04/13/1994 from the Space Shuttle of the Bebedouro region, in Pernambuco, Brazil, band L, polarization HH, pixel = 12.5 meters, 512B 512 pixels, is illustrated in Fig. 8(c). Fig. 8(d) shows a balloon image of the Mojave Desert taken with a CCD camera. Fig. 8(e) shows a Landsat-TM image, band 3, 1000B 1000 pixels, acquired in August/1991 of the Manaus region, in Amazon, Brazil. Fig. 8(f) shows an image of Brasilia, DF, Brazil, with a gallery forest (forest along alluvial areas) and savannah, acquired in 1989 by the SPOT satellite, panchromatic band (resolution 10 meters), 1000B 1000 pixels.

The images were transformed using the similarity model with the following parameters: scale (s), rotation (θ) and translation (Δx and Δy). Table 2 shows the types of images, as well as the distortion model parameters used for image simulation and the parameters estimated by the proposed method. In this table, the measurement units are degree (θ) and pixels (Δx and Δy). One can observe that the estimated parameters are very similar to the real parameters even in Experiment 5 where the rotation is relatively high (θ = 20 degrees). The algorithm parameters were the same for all tests: w = 13, Tn = 0.75 and Tw = 0.9. Values for β were adjusted to generate a sufficiently large number of initial control points, leading to a successful registration procedure. The images were processed in L = 3 (1000B 1000) and in L = 2 (400B 400) resolution levels.

To illustrate the performance of our algorithm, some results obtained in the image simulation and correction processes are presented in Figures 9-11. The figures show the initial control points superimposed on reference and simulated images, with parameters of Table 2, in the lowest level of resolution.

Therefore, to test the registration method and demonstrate its feasibility for real applications, some experiments were performed. An important project at INPE is the study of land use in the Amazonia region (Alves et al., 1994). In this project the image registration task is performed manually with high computational cost. Due to the textural and multi-temporal characteristics of forest images, automatic registration techniques based on closed contours as that one proposed by Li et al. (1995) do not always work properly.

Therefore, the algorithm was tested to register multi-temporal images of forest, agriculture and urban areas obtained by the SPOT-3, Landsat-5 and JERS-1.
satellites (Table 3). The image sizes are 512x512 pixels and they were taken at different times. The algorithm parameters were the same for all tests: $\beta = 3$, $w_c = 13$, $T_c = 0.75$ and $T_w = 0.9$. The images were processed in $\text{L}=2$ resolution levels. Before registration, the radar images were filtered (using a gamma filter) to reduce speckle noise (Oliver and Quegan, 1998).

![Fig. 9 – Initial control points superimposed on (a) reference and (b) simulated images (AES-1) in the lowest level of resolution and (d) Mosaic after registering (\text{L}=2, \beta=1).](image)

![Fig. 10 – Initial control points superimposed on (a) reference and (b) simulated images (Mojav desert) in the lowest level of resolution (\text{L}=2, \beta=1).](image)

![Fig. 11– Initial control points superimposed on (a) original and (b) simulated images (forest) in the lowest level of resolution (\text{L}=2, \beta=2).](image)

<table>
<thead>
<tr>
<th>TABLE 3 – TEST IMAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>TM925AM</td>
</tr>
<tr>
<td>TM945AM</td>
</tr>
<tr>
<td>TM975AM</td>
</tr>
<tr>
<td>TM955AM</td>
</tr>
<tr>
<td>TM965AM</td>
</tr>
<tr>
<td>TM905IT</td>
</tr>
<tr>
<td>TM925IT</td>
</tr>
<tr>
<td>TM9455RIT</td>
</tr>
<tr>
<td>TM955IT</td>
</tr>
<tr>
<td>SP953SP</td>
</tr>
<tr>
<td>SP953DF</td>
</tr>
</tbody>
</table>

Revista Brasileira de Cartografia N° 60/03, outubro 2008. (ISSN 1808-0936) 282
Fig. 12 – Mosaic between original and simulated Landsat images, after correction: (a) reference image (400x400 pixels); (b) simulated image and (d) Mosaic after registering \((L=2, \beta=3)\).

Figures 14-18 show some registration results. Fig. 14 shows the registration of Amazon tropical forest images taken from the TM5 sensor (T2). These images were successfully registered with 485 control points. Fig. 15 shows the registration of two images from the urban area of São Paulo. A SPOT image, band 3 was reduced to 30\text{meter} pixel size and registered with Landsat-TM5 image, band 4. Fig. 16 shows the registration of Amazon region images taken from the TM5 sensor, band 5, in different dates. Fig. 17 shows the registration of Amazon region images taken from JERS-1. Finally, Fig. 18 shows the registration of an urban region images (Brasilia) taken from SPOT and Landsat-TM systems.

Although most of the images used in this work have similar spectral bands, seasonality differences present in these images make it difficult for the automatic registration process. Before registration, images can be preprocessed in order to improve their contrast and to reduce noise.

In order to measure the registration error, we used the root mean square error \(\sqrt{\text{MSE}}\) defined in Equation 23. Control points different from those identified in the registration process (called test points) were manually chosen. Given the spatial transformation and the test points, the error was calculated for each pair.
of images. The results are shown in the last column of Table 3. We find that, in most cases, the registration error is less than one pixel, indicating good registration accuracy.

![Fig. 15](image1.png)

Fig. 15 – Images of urban area: (a) reference image (TM944SP); (b) Registered image (SP953SP); (c) and (d) show initial control points superimposed on reference and warped images in the lowest level of resolution.

![Fig. 16](image2.png)

Fig. 16 – Images of Agriculture area: (a) reference image (TM945I); (b) Registered image (TM905I); (c) and (d) show initial control points superimposed on reference and warped images in the lowest level of resolution.

To investigate the RMSE behavior with different levels of wavelet decomposition the following experiment was performed. A Landsat-TM image of the Amazon region (3354 x 4096 pixels, band 5) was selected and warped by a similarity transformation with parameters $s = 0.90; \theta = 10; \Delta x = 200$ and $\Delta y = 400$.

![Fig. 17](image3.png)

Fig. 17 – Registering JERS93 and JERS96 images: (a) reference image (JERS93); (b) registered image (JERS96); (c) and (d) show the initial control points superimposed on reference and warped images in the lowest level of resolution.

The original and warped images were decomposed by wavelet transforms in 7 levels of resolution ($L = 0, \ldots, 6$). The graph in Fig. 13 shows the refinement of TM95 and TM97 images along the various levels of resolution, starting with $L = 6$ until $L = 0$. At the coarsest level of resolution, the registration transformation is estimated with low precision. As the resolution increases (i.e., as $L$ decreases) the registration transformation is progressively refined until the registration process is completed.

![Fig. 18](image4.png)

Fig. 18 – Images of urban area: (a) reference image (TM944SP); (b) Registered image (SP953SP); (c) and (d) show initial control points superimposed on reference and warped images in the lowest level of resolution.
6. CONCLUSIONS

A multiresolution approach has been successfully applied to two steps of the image registration process: feature identification and matching. The registration method presented here is very simple and easy to apply. A significant amount of computation is saved in comparison to traditional pixel-by-pixel searching methods because the matching is carried out only on the selected feature points and in a coarse-to-fine manner.

Due to the fact that the registration procedure uses the gray level information content of the images in the feature matching process, it is more appropriate to register images of the same sensor or with similar spectral bands. Nevertheless, it has demonstrated good performance for many images of forest, urban and agricultural areas from Landsat-TM and SPOT sensors taken at different times.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Diógenes S. Alves for providing the Amazon region images, Mr. Sergio D. Faria, for providing the agriculture images, José Mura and Luciano Dutra for providing radar images, and finally Dr. Manjunath for providing Mojav Desert images.

REFERENCES


DJAMDIJ, J. P.; BIJAOUI, A.; MANIERE, R. Geometrical registration of images: the multiresolution


KENNEY, C.; MAJUNATH, B. S. Point Matching for Registration with Fit Assessment. To be published. 2001.


