Differential Evolution as a viable tool for satellite image registration

I. De Falco a,*, A. Della Cioppa b, D. Maisto a,c, E. Tarantino a

a Institute of High Performance Computing and Networking, National Research Council of Italy (ICAR-CNR), Via P. Castellino 11, 80131 Naples, Italy
b Natural Computation Lab – DIIE, University of Salerno, Via Ponte Don Melillo 1, 84084 Fisciano (SA), Italy
c School of Graduate Studies, University of Modena and Reggio, Via Campi, 41100 Modena, Italy

Received 31 January 2007; received in revised form 5 September 2007; accepted 21 October 2007
Available online 4 December 2007

Abstract

A software system grounded on Differential Evolution to automatically register multiview and multitemporal images is designed, implemented and tested through a set of 2D satellite images on two problems, i.e. mosaicking and changes in time. Registration is effected by looking for the best affine transformation in terms of maximization of the mutual information between the first image and the transformation of the second one, and no control points are needed in this approach. This method is compared against five widely available tools, and its effectiveness is shown.

#2007 Elsevier B.V. All rights reserved.

Keywords: Differential Evolution; Image registration; Remote sensing; Affine transformation; Mutual information

1. Introduction

Registration is a fundamental task in image processing. During the years several techniques have been developed for various applications resulting in several methods [1,2]. Typically, the registration is a crucial step in the fields of computer vision [3–19] of medical imaging [20–33] and of remote sensing [34–49].

Image registration methods proposed in literature consist of the following four steplike components [1,2]:

- Feature detection. Prominent and distinctive objects (closed-boundary regions, edges, contours, corners, line intersections, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), called control points.
- Feature matching. The correspondence between the features detected in the sensed image and those identified in the reference image is verified. To this aim feature descriptors and similarity measures along with spatial relationships among the features are used.
- Transform model estimation. The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions are calculated by means of the established feature correspondence. According to the transformation model used, mapping functions can be classified into linear transformations which are a combination of translation, rotation, global scaling, shear and perspective components, and elastic or ‘nonrigid’ transformations which allow local warping of image features.
- Image resampling and transformation. The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique.

Among the transformation methods, the one based on the use of an affine transformation [50] to “align” at best the two images to be registered appears of interest in many fields of application. Then the problem becomes that of finding the best among all the possible transformations, each of which is represented by a set of real parameters. An exhaustive search becomes impracticable in case of diverse degrees of freedom of the transformation, and thus heuristic optimization algorithms are helpful. As Evolutionary Algorithms (EAs) [51–54] are successfully applied to face several multivariable optimization tasks, their use has been introduced in image registration as well, in particular in the medical [55–64] and in the remote sensing [65–73] areas.

The goal of the paper consists in the design and implementation of an evolutionary system for the registration of images by using the affine transformation model.

* Corresponding author.
E-mail address: ivanoe.defalco@na.icar.cnr.it (I. De Falco).
Differential Evolution (DE) [53,74] is a version of an EA which has proven fast and reliable in many applications [75–77]. Therefore, we have implemented a DE algorithm to find the optimal combination of the parameter values involved in the affine transformation. There exist in literature several approaches based on either explicitly providing a set of control points [34,70,78–80] (including DE [81]) or in automatically extracting them from the image [82,83,43]. In contrast to those approaches, here we wish to examine DE ability to perform automatic image registration without making any use of control points. This evolutionary system will be tested by means of a set of 2D satellite images.

Paper structure is as follows: Section 2 describes the image registration problem and defines the affine transformation and the mutual information. Section 3 contains DE basic scheme and illustrates the application of our system to the registration task. Section 4 reports on the two remote sensing problems faced, i.e. mosaicking and changes in time, and shows the results achieved by our tool and their comparison against those provided by five widely available registration tools. Finally Section 5 contains conclusions and future works.

2. Image registration

Registration is often necessary for integrating information taken from different sensors (multimodal analysis), or finding changes in images acquired under diverse viewing angles (multiview analysis) or disparate times (multitemporal analysis). Depending on the application, the goals of registering images may be quite different. In remote sensing two problems are usually faced, i.e. Mosaicking and Change Discovery. The former is an example of multiview analysis and deals with spatially aligning two images of neighboring areas taken at the same time so as to obtain a larger view of the surveyed scene, whereas the latter, representing a multitemporal analysis application, consists in firstly aligning two images of about the same area but acquired at different times, and then in pointing out the changes happened in that area within the difference timespan. In all cases, two choices must be made to carry out image registration. The first choice involves the kind of geometric transformation to be considered to find correlations between the given images, while the second one concerns the measure of match (MOM), i.e. the feature on the value of which the goodness of the registration is evaluated. Once made these choices, the MOM can be maximized by using suitable optimization algorithms.

2.1. Affine transformation

The most frequently used transformation model in registration applications is the affine transformation. This model is sufficiently general, since it can handle rotations, translations, scaling and shearing. This transformation can be represented in the most general 3D case as

\[ x' = A \cdot x + b \]  

where \( A \) is a 3 × 3 square matrix accounting for rotations and scalings while \( x, x' \) and \( b \) are three-dimensional arrays representing respectively the original positions, the transformed ones and a translation vector.

2.2. Mutual information

The most widely employed MOM is the mutual information (MI) [84,85], which represents the relative entropy of the two images to be aligned. The greater the value of MI, the better the match between the two images, so this becomes a typical maximization problem.

In general, given two random variables \( Y \) and \( Z \), their MI is

\[ I(Y,Z) = \sum_{y,z} P_{YZ}(y,z) \log \frac{P_{YZ}(y,z)}{P_Y(y) \cdot P_Z(z)} \]  

where \( P_Y(y) \) and \( P_Z(z) \) are the marginal probability mass functions and \( P_{YZ}(y,z) \) is the joint probability mass function. MI is related to entropies by

\[ I(Y,Z) = H(Y) + H(Z) - H(Y,Z) \]  

with \( H(Y,Z) \) being their joint entropy, and \( H(Y), H(Z) \) the entropies of \( Y \) and \( Z \), respectively. The definitions of these entropies are

\[ H(Y) = - \sum_y P_Y(y) \log P_Y(y), \]

\[ H(Z) = - \sum_z P_Z(z) \log P_Z(z) \]

\[ H(Y,Z) = - \sum_{y,z} P_{YZ}(y,z) \log P_{YZ}(y,z) \]

To employ MI as a similarity measure, the 2D histogram of an image pair, the joint histogram \( h \), must be utilized. It is defined as a function of two variables \( Y \) and \( Z \), the gray-level intensities in the two images. Its value at the coordinate \((Y,Z)\) is the number of corresponding pairs having gray-level \( Y \) in the first image and gray-level \( Z \) in the second image. The joint probability mass function of an image pair is then obtained by normalizing the joint histogram of the image pair:

\[ P_{YZ}(y,z) = \frac{h(y,z)}{\sum_{y,z} h(y,z)} \]

From it the two marginal probability mass functions can be obtained as

\[ P_Y(y) = \sum_z P_{YZ}(y,z), \quad P_Z(z) = \sum_y P_{YZ}(y,z) \]

The MI registration criterion states that the image pair is geometrically aligned through a geometric transformation \( T \) when \( I(Y(x), Z(T(x))) \) is maximal. Thus, the aim is to maximize Eq. (3).

The MI methods represent the leading technique in image registration. Depending on the degrees of freedom of the geometric transformation, the search of an optimal similarity measure results in a more or less difficult task, especially if one
considers that small parameter changes may determine meaningful variations in this measure. Several techniques have been proposed to face this problem [86, 87, 84, 88, 89]. Here a Differential Evolution algorithm is taken into account.

3. Differential Evolution

Differential Evolution (DE) is a stochastic, population-based optimization algorithm [53, 74]. It was firstly developed to optimize real parameters of a real-valued function and uses vectors of real numbers as representations of solutions. The seminal idea of DE is that of using vector differences for perturbing the genotype of the individuals in the population. Basically, DE generates new individuals by adding the weighted difference vector between two population members to a third member. This can be seen as a non-uniform crossover that can take child vector parameters from one parent more often than it does from others. If the resulting trial vector yields a better objective function value than a predetermined population member, the newly generated vector replaces the vector with which it was compared. By using components of existing population members to construct trial vectors, recombination efficiently shuffles information about successful combinations, enabling the search for an optimum to focus on the most promising area of solution space.

3.1. DE for Image Registration

3.1.1. Encoding

We have decided to make use of the aforementioned affine transformation model. Since the experiments reported in this paper make reference to couples of two-dimensional images, Eq. (1) reduces to

\[
x'_1 = a_{11}x_1 + a_{12}x_2 + b_1, \quad x'_2 = a_{21}x_1 + a_{22}x_2 + b_2
\]

so the whole problem consists in finding the best combination of six real-valued parameters. Therefore, any individual in the DE population is an array with six positions, with the parameters listed as follows: \( \mathbf{T} = (a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2) \) and each parameter can vary within a range of its own.

3.1.2. Fitness

Given two images \( C \) and \( D \) we take as fitness function their mutual information \( I \), so the aim of the problem becomes to find the best affine transformation \( \mathbf{T} \) for \( D \) such that the mutual information of \( C \) and \( \mathbf{T}(D) \) is maximized.

4. Experiments and findings

We have faced both Mosaicking and Change Discovery problems typical of remotely sensed image registration, as examples of multiview and multitemporal analysis, respectively. The first, named below as Mosaic, accounts for the registration of two images of the same scene acquired from different viewpoints while the second, referred to as Changes, looks for the changes in an area by examining two images taken at different times.

In both cases DE parameters have been set as follows: \( n = 30, g = 200, \text{CR} = 0.5 \) and \( F = 0.5 \). No preliminary tuning phase has been performed. It is important to remark here that, differently from some papers in literature about use of EAs to solve this task, as for example [26, 90, 91], we have decided to use quite wide ranges for each variable in the \( \mathbf{T} \) solution, since we hope that evolution drive the search towards good transformations. The allowed variation ranges are shown in Table 1.

For each problem 20 DE runs have been carried out, so as to investigate the dependence of the results on the initial random

<table>
<thead>
<tr>
<th>Table 1: Problem variable ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_{11} )</td>
</tr>
<tr>
<td>Min.</td>
</tr>
<tr>
<td>Max.</td>
</tr>
</tbody>
</table>
seed. The best of those runs will be discussed below in terms of image transformation achieved and of evolution.

4.1. The Mosaic task

In the first test case we have used two images which are portions of a Landsat Thematic Mapper (TM) digital image recorded on 7 September 1984 over San Francisco bay area (CA, USA) (property of United States Geological Survey [92]). Those images were transformed by us into grey monochannel images, so that each of them is 500 × 500 pixel large and uses 8 bits to represent each pixel. Fig. 1 shows them both. Their $I$ value is 0.1732. Fig. 2 (top left) reports the fusion of the two original images. They share a common area, which should be used by the DE algorithm to find their best registration. Namely, the up-left part of the second image overlaps the bottom-right part of the first, and a slight clockwise rotation was applied to the second image with reference to the first one. So, the best affine transformation should contain a slight counterclockwise rotation and two positive shifts for both the coordinates.

In this problem the best value of $I$ obtained in the best execution is 1.1083. The average of the best final values over the 20 runs is 0.9068 and the variance is 0.1378, the worst result being 0.6351. The best affine transformation found is

$$
x'_1 = 0.946x_1 - 0.253x_2 + 41.858,
$$

$$
x'_2 = 0.253x_1 + 0.946x_2 + 49.779
$$

which represents a counterclockwise rotation of about 15° coupled with a translation in both axes. The resulting transformed image is shown in Fig. 2 (top right). Fig. 2 (bottom left) depicts the fusion of the first original image with the best transformation found for the second one. The alignment of the two registered images is excellent: any detail in the first image, from the streets to the shoreline to the bridges, is perfectly aligned with the corresponding pixels representing it in the transformed second image. In Fig. 2 (bottom right), we report the behavior of the best run achieved for the Mosaic task. In this case, in spite of the very relaxed parameter range allowed, already the initial population achieves an improving solution with respect to the original one. From then on the system proposes many improving affine transformations and both the average and the best fitness values increase over generations until the end of the run.

4.2. The Changes task

In the second test case we have used two images which refer to about the same area but were taken at different times. Namely, they represent an agricultural area near San Francisco (CA, USA) in 1984 and in 1993 respectively (they too are property of USGS [92]). As before, the original Landsat TM images were transformed by us into grey monochannel images, so that each of them is 500 × 500 pixel large with an 8-bit representation for each pixel (see Fig. 3). Their $I$ value is 0.1123. Fig. 4 (top left) reports the fusion of the two original images. As it can be observed, they share a common area, which should be used by the DE algorithm to find their best registration. Namely, the right part of the first image overlaps the left part of the second, and a slight clockwise rotation took place when the second image was taken with reference to the first one. So, the best affine transformation should contain a slight counterclockwise rotation and some shifts for both the coordinates.

In this problem the best value of $I$ attained in the best execution is 0.3951. The average of the best final values over the 20 runs is 0.3918 and the variance is 0.0049, the worst result being 0.3803. The best affine transformation found is

$$
x'_1 = 0.954x_1 - 0.083x_2 + 16.995,
$$

$$
x'_2 = 0.083x_1 + 0.953x_2 + 20.361
$$

which represents a counterclockwise rotation of about 5° coupled with a translation in both axes. The resulting transformed image is shown in Fig. 4 (top right). Fig. 4 (bottom left)
shows the fusion of the first original image with the best transformation found for the second one. The alignment of the two registered images is very good: any detail in the first image, from the rivers to the roads to the fields, is well aligned with the corresponding pixels representing it in the transformed second image. Fig. 4 (bottom right) presents the behavior of the best run achieved for the Changes task. Also in this case, in spite of the very relaxed parameter range allowed, already the initial population achieves an improving solution with respect to the original one and from then on the system proposes many improving affine transformations and both the average and the best fitness values increase over generations until the end of the run.

The computed differences between the first image and the transformed second one are shown in Fig. 5, where only the part in which the two images overlap is meaningful. In it the grey color refers to areas where no changes occurred, the black represents areas that were burned in 1984 and recovered by 1993, whereas the white stands for areas more vegetated in 1984 than in 1993 due to differences in the amount of rainfall, or to the density or level of maturity of the vegetation. Light pixels represent areas burned in 1993 or which were natural landscape areas in 1984 that were converted to agricultural lands and recently tilled, and finally dark pixels stand for areas more vegetated in 1993 than in 1984.

4.3. Comparison and discussion

To investigate the goodness of the results achieved by our tool on the two tasks, we have compared them against those provided by a set of widely available image registration tools. Firstly, we have taken into account ImReg [93], developed at the Vision Research lab at the University of Santa Barbara, USA [94]. A user needs to upload the two images to its internet site and to set a parameter about the desired registration quality (fast, normal, quality, extra). ImReg is based on the following steps: automatical retrieval of a set of tie points, search of the transformation using geometry of tie points, cull of bad tie...
points which do not match this transformation, and finally test of the achieved transformation. We have always used extra quality mode. Since this method has no parameters, only one run can be carried out on any couple of images.

Then we have considered the Image Registration tool (IRT) contained in the Image Processing toolbox of Matlab 7 [95]. The toolbox provides an interactive tool, called the Control Point Selection Tool, that is used by the user to manually pick pairs of corresponding control points in both images. Control points are landmarks that can be found in both images, like a road intersection, or a natural feature. Of course, different sets of control points will determine different transformations. Then, once chosen the kind of transformation (linear conformal, affine, projective, polynomial) IRT will correct the type of distortion present in the base image, determining the parameters of the spatial transformation and transforming the input image to bring it into alignment with the base image. In this case we have chosen an affine transformation and we have performed 20 different runs (corresponding to 20 different choices for the control points) for any pair of images.

Thirdly, we have run a contour-based registration tool again developed at the University of Santa Barbara, USA, and freely downloadable, i.e. Xreg [96]. Its contour matching algorithm is based on the chain-code correlation and other shape similarity criteria such as invariant moments. This approach extracts contour information from each of two images and correlates salient features of the contours: closed contours and the salient segments along the open contours are matched separately. Then Xreg finds optimal transform parameters for aligning the images and registers them. Finally the images are combined to present the result. This method is claimed to work well for image pairs in which the contour information is well preserved, such as the optical images from Landsat and Spot satellites. This algorithm does not depend on external parameters, so only one run has been carried out.

Furthermore, we have used another freely downloadable software tool, RegiStar [97], which is an image alignment, or registration, program that was designed to work specifically with astronomical images. RegiStar finds the stars or other control points in an image, and uses their positions to align this image to another image or group of images. RegiStar uses a sophisticated matching algorithm that allows images at different scales and orientations to be registered and automatically corrects for geometric distortions. It has no parameters, so only one run can be effected.

Finally, we have also tested the capabilities of TurboReg [98], which can be freely downloaded from the site of the Biomedical Imaging Group at the Ecole Polytechnique Federale de Lausanne [99]. It is provided as a plugin for the Image processing and Analysis in Java (ImageJ) tool [100], a public domain Java image processing program developed at the National Institute for Mental Health, USA. TurboReg is an automatic sub-pixel registration algorithm that minimizes the mean square difference of intensities between a pair of 2D or 3D images, uses spline processing, is based on a coarse-to-fine strategy (pyramid approach), and performs minimization according to a new variation of the iterative Marquardt-Levenberg algorithm for non-linear least-square optimization (MLA). The geometric deformation model is an affine transformation. This method is claimed to achieve excellent results in the medical domain. Since this tool is automatic, only one run has been performed.

The results achieved on the Mosaic task are summarized in Table 2. In it we report for any technique the value of the mutual information for the best solution found \( I_B \), and, when meaningful, the average value over the runs \( \bar{I} \), the related variance \( \sigma_I^2 \) and the worst among the best final values found for the mutual information over the 20 runs \( I_w \). Furthermore, for any technique the best solution found is shown. As it can be seen, our DE-based tool achieves the best performance in terms of higher mutual information, and all other methods find solutions with values of \( I \) much lower than that found by DE tool, even in its worst case. Then, TurboReg is in this case the second tool in terms of solution quality.
Similarly, Table 3 contains the results for the Changes task. Also in this case DE turns out to be the best technique, though in this case its superiority with respect to IRT is less impressive. The other methods, instead, keep on being very far from ours and for this problem too the solutions provided by the other methods are much lower than those found by our tool, even in its worst case. It should be remarked here that Registar is not able to carry out the registration task on this couple of images. Moreover it is worth noting that TurboReg shows quite poor performance in this case, although it works well on the Mosaic task. These two latter considerations lead us to suppose that the Changes task is more difficult than expected.

It should be noted that the values of the parameters reported in Tables 2 and 3 are truncated at the third decimal digit. In fact, in some cases, the difference between the value of the same parameter provided by different methods is in the order of $10^{-5}$ to $10^{-6}$. Furthermore, the experimental analysis has proved that small variations in the parameter values may result in

![Image of fusion and transformation results]

**Table 2** Results on Mosaic task

<table>
<thead>
<tr>
<th>Method</th>
<th>$I_b$</th>
<th>$d[l]$</th>
<th>$\sigma_1$</th>
<th>$I_w$</th>
<th>$a_{11}$</th>
<th>$a_{12}$</th>
<th>$a_{21}$</th>
<th>$a_{22}$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>1.1083</td>
<td>0.9068</td>
<td>0.1378</td>
<td>0.6351</td>
<td>0.946</td>
<td>-0.253</td>
<td>0.253</td>
<td>0.946</td>
<td>41.858</td>
<td>49.779</td>
</tr>
<tr>
<td>ImReg</td>
<td>0.3984</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.984</td>
<td>-0.263</td>
<td>0.263</td>
<td>0.984</td>
<td>53.687</td>
<td>37.667</td>
</tr>
<tr>
<td>IRT</td>
<td>0.4719</td>
<td>0.4377</td>
<td>0.365</td>
<td>0.0379</td>
<td>0.944</td>
<td>-0.256</td>
<td>0.256</td>
<td>0.947</td>
<td>49.180</td>
<td>42.365</td>
</tr>
<tr>
<td>Xreg</td>
<td>0.4748</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.966</td>
<td>-0.259</td>
<td>0.259</td>
<td>0.966</td>
<td>36.690</td>
<td>53.647</td>
</tr>
<tr>
<td>RegiStar</td>
<td>0.5342</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.967</td>
<td>-0.256</td>
<td>0.256</td>
<td>0.967</td>
<td>33.051</td>
<td>46.127</td>
</tr>
<tr>
<td>TurboReg</td>
<td>0.5704</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.966</td>
<td>-0.259</td>
<td>0.259</td>
<td>0.966</td>
<td>38.384</td>
<td>45.065</td>
</tr>
</tbody>
</table>
meaningful changes in the value of $I$, thus evidencing that we are dealing with a rugged fitness landscape.

Another important remark which can be drawn by looking at Tables 2 and 3 is that in both problems the variance shown by DE is much lower than that of the other registration tool based on multiple runs, i.e. IRT. This is a very important feature of our algorithm, since implies that it is less dependent on the initial configuration (in this case, the seed for the random number generator) than IRT. Actually, the performance of this latter strongly depends on the choice of the positions of the user-defined control points: the farther one another, the more accurate the resulting registration.

As a conclusion, we have that for both problems our system achieves better solutions than all other methods in terms of higher mutual information, which proves the quality of our approach.

5. Conclusions and future works

In this paper a Differential Evolution strategy has been coupled with affine transformation and Mutual Information maximization to perform automatic registration of remotely sensed images without considering any kind of control points. A comparison has been carried out against five widely available registration tools on two classical problems. The results show that our evolutionary system outperforms the others, and seem to imply that this approach is promising, yet there is plenty of work still to do. Therefore, future works shall aim to evaluate the effectiveness of our system in this field, and its limitations as well.

Firstly, a wide tuning phase shall be carried out to investigate if some DE parameter settings are, on average, more useful than others. Moreover, we aim to apply our approach also to multimodal analysis in remote sensing, as for example fusion of information from sensors of different characteristics. Furthermore we plan to implement a coarse-grained parallel version of the DE algorithm based on the island model, and to run it on a cluster of workstations.

Lastly, our final goal is to design and implement a technique which could be useful also for 3D multimodal medical image registration.

References


