An Evolution Strategies Based Approach to Image Registration

Abstract

An image registration approach based on Evolution Strategies is proposed. In image registration, an invariant reference needs to be established within each source image, which is unavailable in many cases. To solve this problem, feature configuration (which is defined as the cluster of feature vectors on an image representing homogeneous feature distribution,) is employed to describe the object inside the scene. Instead of finding the correspondence of the entire image, the spatial relationships of the feature configuration in every source image are discovered with Evolution Strategies (ES). While one approach, even this one, may not be suitable for every image domain, the ES approach has many advantages. Compared to some methods, it is computationally effective; compared to other, it is capable of discovering transformations of larger scope (e.g., greater rotation angles or translation distances etc.) The search structure we use is an ellipsoid. The results from various images prove it to be an efficient and effective method.

1 INTRODUCTION

A fundamental image-processing task, image registration matches two or more images such that features from each individual source are aligned against the same reference. Virtually all image understanding tasks, such as image fusion, object recognition etc., require image registration procedure as pre-processing. It is a particular important issue faced in almost all remote sensing domains. In medical imaging, a patient’ s cranial scan must be matched with medical atlas images as well as previous scans of the same patient. In Earth science, the extent of deforestation can be determined only if the present image can be compared to ones from previous time periods. These and many other examples exist that assert the need to put multiple images into pixel-by-pixel correspondence.

Intensive research has been devoted to find the most effective and efficient registration methods [2, 8, 14, 17]. Approaches proposed include control point based methods [13], frequency feature based methods [1, 16], mutual information based methods [9] etc. Figure 1 illustrates an image registration example. Two images are superimposed one on the other, as shown in Figure 1(c), based on the transformation discovered by the registration process.

Figure 1. Image (a) and (b) are two source images focusing on the same object but containing different scene. (c) is the superimposed image after registration.

Image registration can be viewed as a combination of different choices of feature space, search space, searching strategy, and similarity measure [12]. The feature space extracts information from the source images, which provides a quantitative space for transformation. All possible transformations form a search space, such that given a pair of images a sequence of transformations can be found in the search space to put these images in correspondence. The searching strategy defines rules of
finding the next transformation. The registration accuracy is assessed by the similarity measure. The registration process proceeds iteratively by searching and applying transforms until the similarity measurement is satisfied.

Generally, a similarity measure is required to evaluate the accuracy of aligned image after transformation. During registration, the similarity measurement is iteratively computed and improved by adjusting the transformation. However, most similarity measurements are computationally expensive even on moderately sized images, which makes registration inefficient, especially when there is a large initial difference between images. Here we propose a fast, control point free and feature based image registration method, which is based on Evolution Strategies. The objective is to find the correspondence between two or more images having a spatial difference caused by rigid transformation.

The principal idea is to search for the correspondence between some specific feature configurations instead of the correspondence across the whole image. Given the specific feature description, Evolution Strategies is employed to find the feature configuration, defined in section 3.2, on all source images. A reference image is randomly selected thereafter, in which the feature configuration is located and is used for registering other images. However, the region defined by feature configurations may not enclose the same feature distribution. Therefore, a refinement process based on the reference image is followed to adjust the feature configuration such that similar feature enclosure is ensured. Finally, the transformation functions are determined by comparing the spatial characteristic of configurations, which is represented in the form of feature ellipse, against that of the reference configuration and are used to register those images.

The rest of this article is organized as follows. In section 2, a short review of Evolution Strategies is given. Section 3 presents the retrospective image registration problem and illustrates Evolution Strategies based image registration scheme. Experiments are illustrated in section 4. The paper is concluded with discussion in section 5.

2 ES SHORT REVIEW

Evolution Strategies (ESs) are algorithms that imitate the principles of natural evolution as a method to solve parameter optimization problem [4, 3, 18]

The goal of a parameter optimization problem \( f: \mathbb{R}^n \rightarrow \mathbb{R}, \ \mathbb{R}e \theta \), where \( f \) is called objective function, is to find a vector \( x \in M \) such that

\[
\forall x \in M : f(x) \geq f(x')
\]

(1)

where \( f' := f(x') \rightarrow \infty \) is called a global minimum; \( x' \) is the global minimum point. \( M \) is the set of feasible points for a problem. In correspondence with global minimum, a local minimum \( f = f(\hat{x}) \) is defined by the following condition:

\[
\exists \varepsilon > 0 \ \forall x \in M : \| x - \hat{x} \| < \varepsilon \Rightarrow f = f(\hat{x})
\]

(2)

Coexistence of global minimum and several local minima makes optimization a non-trivial problem.

Each ES individual represents a vector within the domain of the objective function \( f \). Each \( x_i, i = 1, 2, ..., n \), is termed an object variable and is represented as a real value in the individual. Evolution Strategies is essentially randomized hill climbing, which makes it a non-deterministic optimization strategy. Hill climbing necessitates the resolution of two issues at each iteration -- (i) the direction to move and (ii) the distance (step size). These issues are resolved by embedding control variables into individual and an ES individual is, organized as object variables and control variables, illustrated below.

\[
\{ x_1, x_2, ..., x_n, \sigma_1, \sigma_2, ..., \sigma_n, \theta_1, ..., \theta_p \}
\]

(3)

where the \( \sigma \)'s and \( \theta \)'s are control variables.

An object variable should be considered the mean of a normally distributed random variable. Under that interpretation, each \( \sigma_i \) is a standard deviation for an object variable. Thus \( m \leq l \) (if \( m < l \) then \( \sigma_i \) applies to all \( x_i, m \leq j \leq l \) ). Each \( \theta \) is a surrogate for the covariance of two object variables. \( \Theta \) is organized as an upper triangular matrix as a covariance matrix. (That is to say, \( p = (2l-m)(m-l)/2 \).) The correspondence between \( \Theta_{ij} \), \( i \leq j \leq l \) and the covariance, \( c_{ij} \) is

\[
\tan(2\theta_{ij}) = \frac{2c_{ij}}{\sigma_i \sigma_j}
\]

(4)

An interpretation of an ES organism is an \( l \)-dimensional jointly distributed normal variate with mean \( x \) and the standard deviation \( \sigma \). The orientation of the distribution in \( l \)-space is determined indirectly by the covariance and directly by \( \Theta \).

The incorporation of control variables into the individual representation establishes a two-level self-learning process, since not only the object variable adapts according to the objective function, but also the control variables change with respect to the actual topological requirements. In other words, the control variables make up an internal model of the objective function, which is learned on-line during optimum seeking without an additional measure of fitness.

The ES algorithm is formulated in the language of biology as follows:

Step 1. A given population consists of \( \mu \) individuals. Each is characterized by its genotype consisting of \( n \) genes, which unambiguously determine the fitness for survival.

Step 2. By mutation and recombination operations, each individual parent produces \( \lambda \mu \) offspring on average, so that a total number of \( \lambda \) offspring individuals are available.

Step 3. Select the best of the offspring to form parents of the following generation and continue at Step 1.
3 IMAGE REGISTRATION WITH ES

Let \( I_1 \) and \( I_2 \) denote two image matrices, then image registration, under Cartesian Coordinates, can be expressed as:

\[
I_1(x, y) = g(I_2(f(x, y)))
\]

where function \( f(.) \) is a 2D spatial-coordinate transformation, i.e. \( f(.) \) maps two spatial coordinates, \( x \) and \( y \), to new spatial coordinates \( x' \) and \( y' \), and function \( g(.) \) is a 1-D intensity or radiometric transformation.

The registration problem is to find the optimal transform functions \( f(.) \) and \( g(.) \), namely spatial and intensity transformation, so that the images are aligned under the same coordinates system. The intensity transformation \( g(.) \) is not always necessary, and if \( g(.) \) is need, a lookup table is usually sufficient [12].

3.1 RETROSPECTIVE REGISTRATION

Retrospective registration is required when an image is obtained without the benefit of a fiducial reference system, e.g. a battle field surveillance image or MRI cranial scan image. In this case, a reference is not included in the source image.

Let point \((x, y)\) denotes the central point of the object and let the rotation angle be \( \Theta \). In order to describe the scaling transformation, let’s denote a point in an image in homogeneous coordinates \((x, y, s)\), with \((x/s, y/s)\) being the transformation, let’s denote a point in an image in homogeneous coordinates \((x, y, s)\), with \((x/s, y/s)\) being the transformation. Using the homogeneous coordinates, the transformation function \(f(x, y)\) has the following general form:

\[
f_{\Theta, t_x, t_y}(x, y, s) = \begin{bmatrix} \cos(\Theta) & -\sin(\Theta) & t_x \hline \sin(\Theta) & \cos(\Theta) & t_y \hline 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ s \end{bmatrix}
\]

where \( \Theta \) determines the rotation angle. \( t_x/s \) and \( t_y/s \) are translations on horizontal and vertical directions. \( s \) is the scaling weight that also functions as the normalization factor in the homogeneous coordinates.

3.2 REGISTRATION WITH ES

To deal with retrospective registration, an image feature is employed in our scheme. Frequently used features include luminosity, texture, shape, etc. Usually, a picture contains at least one object, which can be distinguished from its background with a set of characteristic features. Such features are distributed in a certain region, i.e., the area defined by the shape of that object. That is, the co-occurrence of these features is only satisfied within the object. Hence, the area that concurrently contains a set of certain features is called the feature configuration. The feature configuration, \( \Phi \), is defined on image \( I \) such that a cluster of pixels constitute a close region \( P \) within which image features, \( F_1, F_2, ..., F_n \), are uniformly distributed.

\[
\forall(x, y), (x', y') \in P \quad \text{and} \quad \| I_1^{F_1, F_2, ..., F_n}(x, y) - I_1^{F_1, F_2, ..., F_n}(x', y') \| \leq \varepsilon, \quad \varepsilon \rightarrow 0
\]

and \( \Phi_{F_1, F_2, ..., F_n}(x, y) = \{ I(x, y) \mid I(x, y) \in P, P \subseteq I^{F_1, F_2, ..., F_n} \} \)

where \( I_1^{F_1, F_2, ..., F_n}(x, y) \) is the feature vector generated by applying feature filters onto image \( I \) and function \( | \cdot | \) measures the distance of two vectors.

The rigid transformation does not change the feature of an image. That is, given the initial image \( I_1 \) and the transformation function \( f \), the outcome image \( I_2 \) has the same feature as \( I_1 \). Thus, the feature configurations of these features in \( I_1 \) and \( I_2 \) are also spatially related with the same transformation.

\[
\Phi_{F_1, F_2, ..., F_n}(x, y) = f(\Phi_{F_1, F_2, ..., F_n}(x, y))
\]

Because of its separability, a transformation function can also be expressed as follows:

\[
I_2(x, y) = S(T(I_1(x, y), \Delta x, \Delta y), \Delta \Theta), s)
\]

where \( T(\cdot), R(\cdot) \) and \( S(\cdot) \) denote translation, rotation and scaling respectively. Notice that the translation parameters \( \Delta x \) and \( \Delta y \) are defined with regard to the central point of the image. To prove equation (8), it is sufficient to prove that after translation and rotation, the distance between any two points in the feature configuration remains unchanged, while scaling enlarges the distance by scale of \( s \). Due to space constraints, proofs are not included.

Yuan et al. reported successful feature identification using Evolution Strategies [18]. Inspired by the success of their work, Evolution Strategies is employed to search for the optimal transformation. During the search for image correspondence, instead of evaluating similarity measurements over the whole image, ES identifies the region from each source image that contains a homogeneous feature configuration. To capture the feature configuration, an ellipse structure is used to enclose the maximum homogeneous feature area. Here we call such an ellipse the feature ellipse.

Feature Ellipse

Feature ellipse, \( \Lambda \) is the search structure used to enclose a feature configuration \( \Phi_{F_1, F_2, ..., F_n} \). Feature vectors inside the ellipse represent the same type of features. That is, it encloses only one type of feature configuration.

\[
\Lambda(x_0, y_0, a_1, a_2, \Theta) \mid \Phi_{F_1, F_2, ..., F_n}
\]

\[
\Rightarrow | I_1^{F_1, F_2, ..., F_n}(x, y) - I_1^{F_1, F_2, ..., F_n}(x', y') | \leq \varepsilon
\]

and \( \varepsilon \rightarrow 0 \)

where \( I_1^{F_1, F_2, ..., F_n}(x, y) \) and \( I_1^{F_1, F_2, ..., F_n}(x', y') \) are any two different feature vectors enclosed by \( \Lambda \). A feature ellipse is determined by the coordinates of the center \((x_0, y_0)\), the lengths of the major and minor axes \(a_1\) and \(a_2\), and the angle \( \Theta \) between the major axis and the horizontal line. These parameters are embedded into ES as objective variables and are organized as \((x_0, y_0, a_1, a_2, \Theta)\). Figure 2 illustrates the structure of a feature ellipse. (Notice that ES also uses control variables denoted as \( \Theta \) s that
establish a relationship with covariance as illustrated in equation (2) and directly determine the orientation of the distribution in $l$-space.

The optimum to the objective function of one ES application is a feature ellipse that maximizes its area under the constraint that only one feature configuration is included. Or in the other words, it minimizes the difference among the feature vectors inside the ellipse while enlarging its area.

**Parameters Estimation**

Given the feature description, ES is applied to all source images to find the feature configuration $F_i$, $i=1,...,n$. The feature configuration $F_i$ located from one of the source images, which is randomly selected, is used as the reference for registering other images. This arbitrarily selected image is distinguished as the reference image. Figure 3 illustrates the diagram of ES based image registration.

Although the same feature description is used to guide the ES search, the outcome $F_i$ usually does not give the same quantitative measurement, which appears as slightly different feature ellipses as shown in Figure 4(c) and 4(d). This is because of the probability-controlled randomness of reproduction process in ES. Therefore, a refinement process is followed. The refinement takes $F_i$ as a reference and adjusts the feature configuration $F_i$ of image $I_i$, such that the quantitative measurements, e.g. mean and variance, of each feature configuration, $F_i$ and $F_i'$, match.

After searching, the feature configuration is reported as the parameters of an ellipse. The transformations $f_i$, $i=2, \ldots, n$, are determined by comparing the spatial parameters of ellipses with that of the reference.

**4 EXPERIMENTS**

In our experiments, the parent population size is chosen as 50 and the descendant population size is 300. These population sizes are used in both the initial search as well as the later refinement step. For the purpose of accelerating the search process, discrete recombination on object variables and panmictic intermediate recombination of control variables is preferred [21, 3, 22].

Moments are used in our experiments as the quantitative measurements of the feature configuration. Generally, only the first few moments are required to differentiate between signatures of clearly distinct shapes [7, 6].

$$
\mu_i(v) = \sum (v_i - m)^n p(v_i)
$$

The quantity $m$ is recognized as the mean of $v$, which represents the gray level of an image, and $\mu_i$ as its variance. $p(v_i)$ is the normalized amplitude histogram at gray level $v_i$. Besides the moments, the compactness of the ellipse, which is the ratio of the major axis and the minor axis, is considered as another constraint of the feature ellipse. Given the match of the compactness measurement, the ratio of the axes from two feature ellipses exposes the scale factor of the image. Let the axes of two feature ellipses be $(a_1, a_2)$ and $(a_1', a_2')$ where $a_1, a_1'$ are the length of major axes and $a_2, a_2'$ are the length of minor axes. The scaling factor is determined by the average of the ratio as shown below.

$$
s = \frac{a_1/a_1' + a_2/a_2'}{2}
$$

Figure 4 illustrates the process with sample images Child_1 and Child_2. An initial feature ellipse is randomly chosen for each source image, which are drawn and shown in Figure 4(a) and 4(b). After approximately 50 generations, both feature configurations are found, as shown in Figure 4(c) and 4(d).
Although two ES processes are running under the same feature description, regions covered by two feature ellipses are usually not identical. This is due to the non-deterministic search characteristic of the ES algorithm. Therefore, a refinement process is followed. In this experiment, Child_1 is chosen as the reference image. The mean and variance values of the area inside feature ellipse in image Child_1 are computed and used to refine the parameters of feature ellipse in Child_2. The refinement process uses ES optimization with one more constraint. That is, the difference between the quality measure of feature ellipse in Child_2 and the quality measure extracted from Child_1 is minimized and allow 1% error for mean, 0.1% error for variance and 1% error for the ellipse compactness.

Table 1 lists the estimated transformation on six pairs of test images. Each pair of images is of the same size. The first two columns give the condition and the name of images used in the experiments. The next three columns contain the Rotation and Translation parameters. The noisy images are distorted with 20% noise.

<table>
<thead>
<tr>
<th>Test Image</th>
<th>Rotation</th>
<th>Translation</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Image 1</td>
<td>Image 2</td>
<td>Image 1</td>
</tr>
<tr>
<td></td>
<td>Δθ</td>
<td>(Δx, Δy)</td>
<td>(Δx, Δy)</td>
</tr>
<tr>
<td>Noise Free</td>
<td>Child</td>
<td>-0.3</td>
<td>-9.4</td>
</tr>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plane</td>
<td>-2.1</td>
<td>28.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td>44.5</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>River</td>
<td>21.6</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infrared</td>
<td>-5.1</td>
<td>-31.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lighthouse</td>
<td>86.0</td>
<td>75.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noisy Image</td>
<td>Child</td>
<td>-0.3</td>
<td>-9.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plane</td>
<td>-0.3</td>
<td>29.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td>45.3</td>
<td>75.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>River</td>
<td>24.4</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infrared</td>
<td>-6.5</td>
<td>-33.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lighthouse</td>
<td>-3.6</td>
<td>-12.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
estimated rotation parameters, where results in column ‘Image 1’ are the principal orientation of the feature ellipse located in the first image and column ‘Image 2’ contains orientation parameters computed from the second image. The column $\Delta \theta$ lists the rotation difference between image pairs. The last three columns contain the translation estimations, which are coordinates of the central points of ellipses found in image pairs and the translation difference, $\Delta x$ and $\Delta y$, between central points. Table 1 also contains the outcomes performed on the same experimental images, except each image is distorted with 20% noise. It is clear that the transformation parameters estimated under noise are very close to those computed with noise-free images.

Figure 5 illustrates two registered images, lighthouse and Child, given the transformation parameters provided in table 1. The registrations are accurate under the judgment of human perspective.

Figure 5. Sample registration results, lighthouse and child. Two source images are superimposed one on the other using the transformation matrix estimated with ES optimization.

Notice, the translation parameters $\Delta x$ and $\Delta y$ are not based on the center of the image but the central point of the feature ellipse. Therefore, when registering images, the translation and rotation are performed with regard to the center of the ellipse.

Table 2 illustrates the experimental result on scaling factor estimation. Notice that test images are resized to 80% of the original size. The second and the third columns are the axes’ lengths of the feature ellipses, which are formatted as (length of major axis, length of minor axis). The error estimation is below 0.02.

<table>
<thead>
<tr>
<th>Images</th>
<th>Axes (long, short)</th>
<th>Scale Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100% image</td>
<td>80% image</td>
</tr>
<tr>
<td>Child</td>
<td>(197.0, 83.3)</td>
<td>(172.0, 62.7)</td>
</tr>
<tr>
<td>Plane</td>
<td>(123.7, 38.4)</td>
<td>(95.3, 32.5)</td>
</tr>
<tr>
<td>Phone</td>
<td>(130.0, 55.7)</td>
<td>(103.9, 44.3)</td>
</tr>
<tr>
<td>River</td>
<td>(120.3, 24.7)</td>
<td>(93.7, 20.4)</td>
</tr>
<tr>
<td>Infrared</td>
<td>(181.8, 60.7)</td>
<td>(148.2, 47.1)</td>
</tr>
<tr>
<td>Lighthouse</td>
<td>(114.9, 50.6)</td>
<td>(91.4, 40.9)</td>
</tr>
</tbody>
</table>

5 DISCUSSION AND CONCLUSION

An Evolution Strategies based image registration approach is described in this article. Given the feature of an image remaining unchanged after rigid transformation, a search structure, the feature ellipse, is embedded into each ES individual. The objective function of ES achieves the minimization of differences between the quantitative measurements of feature configurations enclosed with feature ellipses while enlarging its area. The registration scheme contains three steps: search feature configurations under certain feature description, refining feature ellipses by minimizing quantitative measurements differences, and determining the transformation parameters.

Experiments have been performed on various kinds of images including nature scenes, military surveillance images etc. Promising results are also obtained under noisy circumstances. The experiments show the robustness of this approach, which is the result of two aspects. Firstly, the search is performed in the feature space, where the noise is reduced. Secondly, the optimization process, incorporating with the feature quantitative measurements, is insensitive to the noise. Even though the feature configurations found within the noise-free images and within noisy images are different, the transformation relationships discovered are almost identical.

Moreover, since the feature comparison is performed inside a relatively small region, the feature ellipse, the computation expense is reduced. Figure 6 illustrates that the ratio of variance vs. area (enclosed by feature ellipse) changes with regard to the iterations. Figure 6(a) illustrates the optimization process with a Phone image. In the graph, solid circle line records the ratio changes of the reference image. The solid triangle line records the ratio changes of the companion image. (Remember, the reference image is simply an arbitrarily selected source image.) The graph is partitioned with a vertical dotted line, where the left half illustrates the searching phase optimization and right half shows the refinement phase progress. Notice that during the first phase of registration, ES is applied individually on each image. Therefore the iterations used in searching are
different. In figure 6(b), which illustrates the optimization process with a Plane image, it is clear that the refinement process adjusts the feature ellipse in the companion image even closer to that of the reference image.

Experiments have been successful on images containing one object, which is distinguishable with a set of features. In cases where image contains more than one similar objects, due to the non-deterministic characteristic of ES, feature configurations representing different but similar objects may be found, e.g., the image shown in Figure 7. Obviously, it is hard to further distinguish among these objects. This difficulty may be overcome by ES with a niching strategy. Zhang et al. [19] described a niching embedded ES for multimodal function optimization, in which successful locating multiple optima is reported. Further study can be done for multi-object involved registration.

Figure 6. Variance reduces during searching and refinement. Axis X is the iteration number and axis Y is the ratio of variance vs. area enclosed within feature ellipse. In each graph, reference image is illustrated with solid circle line and the companion image is illustrated with solid triangle line.

Figure 7. Confusion of multiple similar objects in image registration. Two similar feature configurations are found yet corresponding to different objects.

Acknowledgments

Support for this work was provided in part by DoD EPSCoR and the Board of Regents of the State of Louisiana under grant F49620-98-1-0351.

References


