

Automated Image Registration Using Morphological Region of Interest Feature Extraction

Antonio Plaza¹, Jacqueline Le Moigne² and Nathan S. Netanyahu^{3,4}

¹Department of Computer Science, University of Extremadura, Avda. de la Universidad s/n, 10071 Caceres, Spain

²Computational and Information Sciences and Technology Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771

³Department of Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel

⁴Center for Automation Research, University of Maryland, College Park, MD 20742

Abstract—With the recent explosion in the amount of remotely sensed imagery and the corresponding interest in temporal change detection and modeling, image registration has become increasingly important as a necessary first step in the integration of multi-temporal and multi-sensor data for applications such as the analysis of seasonal and annual global climate changes, as well as land use/cover changes. The task of image registration can be divided into two major components: (1) the extraction of control points or features from images; and (2) the search among the extracted features for the matching pairs that represent the same feature in the images to be matched. Manual control feature extraction can be subjective and extremely time consuming, and often results in few usable points. Automated feature extraction is a solution to this problem, where desired target features are invariant, and represent evenly distributed landmarks such as edges, corners and line intersections. In this paper, we develop a novel automated registration approach based on the following steps. First, a mathematical morphology (MM)-based method is used to obtain a scale-orientation morphological profile at each image pixel. Next, a spectral dissimilarity metric such as the spectral information divergence is applied for automated extraction of landmark chips, followed by an initial approximate matching. This initial condition is then refined using a hierarchical robust feature matching (RFM) procedure. Experimental results reveal that the proposed registration technique offers a robust solution in the presence of seasonal changes and other interfering factors.

Keywords—Automated image registration, multi-temporal imagery, mathematical morphology, robust feature matching.

I. INTRODUCTION

In Earth Science, Space Science and soon in Exploration Science, many current and future space applications are, or will be requiring real- or near-real time integration of data acquired from multiple sources. For real-time applications, multiple source data integration provides the necessary information to intelligent navigation and decision support systems. Depending on the application at hand, this integration may be performed on-board or on-the-ground, and in the case of Earth Science, these multiple sources may be satellite, aircraft and ground measurement systems and the data represent multiple temporal, spectral and spatial resolutions. For such multiple source integration, a fast, automatic, reliable and accurate image registration is an essential step.

Assuming that image data have been systematically corrected through a prior navigation process, image registration or “precision correction” corresponds to a feature-based matching of the image data and enables to refine the geo-registration to a sub-pixel accuracy. Image registration is then defined as the process that determines the best match between two or more images acquired at the same or at different times, by different or identical sensors. One set of data is taken as the reference data and all other data, called input data (or sensed data), are matched relative to the reference data. The general process of image registration includes three main steps: (1) the extraction of features to be used in the matching process, (2) the application of a feature matching strategy subject to a specific metric, and (3) the re-sampling of the data based on the correspondence computed from matched features. A large number of automatic image registration methods have been proposed and surveys can be found in [1], [2]. Our previous work [3] has focused on the comparison of different choices for steps (1) and (2), where either an entire scene was being registered (against a reference scene), or where several (small) chips, extracted typically around characteristic image features or landmarks from the reference scene, were being registered initially against corresponding windows from the input scene. The working assumption was that once a registration system became operational, a database of such landmark chips would be available for use. Developing, however, the capability of extracting these chips automatically and independently of any database – for example, through the use of mathematical morphology (MM) techniques [4], [5] – presents the following advantages:

- It eliminates the need of maintaining a chip database that would have to be updated regularly and adapted to each sensor.
- If reference and input data are pre-processed with a cloud detector, an automatic chip extractor can take into account the cloud mask and therefore automatically find chips/windows containing a very small amount of clouds.
- It enables any algorithm to process any size images.
- Extracting simultaneously corresponding chips from the reference image with windows from the input image brings the initial registration error down,

enabling feature matching methods (or any optimization strategy) to work better and faster, since the initial conditions are closer to the final transformation, i.e. these methods serve as "refinement strategies."

In this paper, we propose to add the following initial step (0) to the previously described image registration process: (0) the extraction of reference chips and corresponding input windows on which local registrations are performed. The remaining of this paper describes how step (0) can be performed using MM. Furthermore, we also show how performing multiple local registrations requires a post-processing step (at the end of step (2)) that computes a more accurate global transformation.

The paper is organized as follows. Section II shows an overview of classic MM, including basic MM operations. Section III describes a MM-based region of interest feature extraction algorithm, used in this work as step (0) of a fully automated image registration system. This algorithm makes use of a spectral dissimilarity metric (spectral information divergence) to obtain a first approximation matching, which is then refined using a robust feature matching (RFM) procedure outlined in Section IV. Section V presents empirical results obtained by applying our automated registration scheme to multi-temporal Landsat scenes collected over Central Virginia. Finally, Section VI concludes with some remarks and hints at plausible future research.

II. CLASSIC MATHEMATICAL MORPHOLOGY

Mathematical morphology (MM) is a non-linear image processing technique that was originally established by introducing fundamental operators applied to two sets [4]. One set is processed by another one having a carefully selected shape and size, known as the structuring element (SE), which is translated over the image. The SE acts as a probe for extracting or suppressing specific image structures by checking that each position of the SE fits within the image objects. Based on these notions, two fundamental operators are defined in classic MM, namely *erosion* and *dilation*. The application of the erosion operator to an image gives an output image, which shows where the SE *fits* the objects in the image. In contrast, the application of the dilation operator to an image gives an output image, which shows where the SE *hits* the objects in the image. All other MM operations can be expressed in terms of erosion and dilation. For instance, the notion behind the *opening* operator is to dilate an eroded image in order to recover as much as possible of the eroded image. In contrast, the *closing* operator erodes a dilated image so as to recover the initial shape of image structures that have been dilated.

In grayscale morphology, each 2-D gray tone image is viewed as if it were a digital elevation model (DEM). Following a usual notation [5], let us consider a grayscale image f , defined on the 2-D discrete space Z^2 , and a SE designed by $B \subset Z^2$. The latter is usually "flat" in the sense

that it is defined in the spatial domain of the image (the $x-y$ plane). The erosion of f by B is defined by:

$$(f \otimes B)(x, y) = \bigwedge_{(s,t) \in Z^2(B)} f(x+s, y+t), \quad (x, y) \in Z^2, \quad (1)$$

where $Z^2(B)$ denotes the set of discrete spatial coordinates associated to pixels lying within the neighborhood defined by B , and \bigwedge denotes the minimum. In contrast, the dilation of f by B is defined by:

$$(f \oplus B)(x, y) = \bigvee_{(s,t) \in Z^2(B)} f(x-s, y-t), \quad (x, y) \in Z^2, \quad (2)$$

where \bigvee denotes the maximum. Using the same notation above, standard opening and closing filters [5] can be defined respectively by $(f \circ B)(x, y) = [(f \otimes B) \oplus B](x, y)$ and $(f \bullet B)(x, y) = [(f \oplus B) \otimes B](x, y)$. The two operations above have been successfully employed in a variety of geoscience and remote sensing applications [6]. In particular, a composition of opening and closing operations using SEs of different sizes was used in [7] in order to build pixel-level scale-based morphological profiles, which were then used to characterize image structures in urban satellite data. In certain applications, however, orientation is a worthy addition to scale in order to characterize image features.

III. MORPHOLOGICAL REGION OF INTEREST EXTRACTION

A. Scale-Orientation Morphological Profiles

Scale-orientation morphological profiles (SOMPs) are obtained using classic MM concepts by applying openings and closings with line segment SEs of varying orientation [8]. In order to define the SOMP concept in mathematical terms, we first denote by $B_{p,(dx,dy)}$ a line segment SE with minimal length, where p is the number of pixels along the line and dy/dx is the slope of the line segment [9]. By assuming a management of images digitized on a square grid, we can restrict our analysis to line slopes in the form of an irreducible fraction dy/dx . By convention, we include the forms $0/1$ and $1/0$ in referring to horizontal and vertical lines respectively. Let us assume that: (1) A basis set contains a collection of (dx, dy) pairs which define the orientations of the considered line segment SE, and (2) The resulting line segment SEs are approximated on a discrete grid depending on the length of the line segment. With the above assumptions in mind, we define the scale-orientation morphological profile (SOMP) by opening at a given pixel (x, y) of an image f as:

$$D_{(dx,dy)}^{\circ}(x, y) = \left\{ \left[(f \circ B_{p,(dx,dy)}) \bullet B_{p,(dx,dy)} \right] (x, y) \right\}, \quad (3)$$

Similarly, we define the SOMP by closing at the pixel $f(x, y)$ as:

$$D_{(dx, dy)}^{\circ}(x, y) = \left\{ \left[(f \bullet B_{p,(dx, dy)}) \circ B_{p,(dx, dy)} \right] (x, y) \right\}, \quad (4)$$

where $p = \{1, 2, \dots, k\}$ and k is the maximum number of considered scales. In both cases, a measure of line strength can be computed for each scale and orientation by calculating the Euclidean distance (ED) between pixel (x, y) in the original image and the pixel at the same location in the image filtered by the considered line segment SE as follows:

$$s^{\circ}(x, y) = \left\{ \text{ED} \left(f(x, y), \left[(f \circ B_{p,(dx, dy)}) \bullet B_{p,(dx, dy)} \right] (x, y) \right) \right\} \quad (5)$$

$$s^{\bullet}(x, y) = \left\{ \text{ED} \left(f(x, y), \left[(f \bullet B_{p,(dx, dy)}) \circ B_{p,(dx, dy)} \right] (x, y) \right) \right\} \quad (6)$$

The resulting values are combined in a feature vector $D(x, y) = s^{\circ}(x, y) \cup f(x, y) \cup s^{\bullet}(x, y)$ with dimensionality $2k \times m$, where l is the number of considered orientations. In order to automatically extract significant points of interest for registration, we use the concept of self-information as a measure of dissimilarity [10] (the more irregular the SOMP, the higher the chances that it represent a pixel in the corner of an object). Let us denote by $d_i(x, y)$ each of the components of $D(x, y)$. Since $D(x, y)$ can be considered as an information source, we can further define its self-information provided by the n -th component as $I_n(x, y) = -\log p_n(x, y)$, where $p_n(x, y) = d_n(x, y) / \sum_n d_i(x, y)$. Using the above definitions, the entropy of $D(x, y)$ can be obtained as follows:

$$H(D(x, y)) = -\sum_n p_i(x, y) \cdot \log p_i(x, y) \quad (7)$$

B. Automated SOMP-Based First Approximation Algorithm

With the ultimate goal of integrating automatic registration into a processing system for multi-temporal data, we propose a morphological algorithmic framework that can be described by the following main steps. The input to the algorithm consists of a reference scene, f , from which a set of landmark chips will be automatically extracted, and an input scene, g , where corresponding windows are found in an unsupervised manner, a chip (and window) size, and a size of a subimage in which to search for a window (see below for specific details). These chip-window pairs will help obtain an initial condition, i.e., a first-cut approximation for an automated registration framework. The algorithm is given by the following steps:

1. For each pixel (x, y) in the reference scene, f , construct a SOMP, denoted by $D_f(x, y)$, and calculate its entropy, $H(D_f(x, y))$.
2. Select the pixel $f(x', y')$ with the highest $H(D_f(x, y))$ score (i.e., with the maximum entropy) in the reference scene.
3. Extract a reference chip centered on the considered pixel with maximum entropy, (x', y') . The size of the reference chip is specified by an input parameter (e.g., 256x256 in our experiments).
4. Obtain the SOMP for each pixel in the input scene, i.e. compute $D_g(x, y)$ for each (x, y) . In this work, since the input scenes are geometrically and radiometrically corrected, a local search area around pixel coordinates (x', y') in the input scene is considered for finding a pixel whose SOMP best matches $D_f(x', y')$ (see next step). The size of the local search area is specified as an input parameter (e.g., 1000x1000 in our experiments).
5. For each pixel (x, y) in the local search area of the input scene, compute the root mean square error (RMSE) between the SOMPs $D_g(x, y)$ and $D_f(x', y')$. Retain the pixel coordinates, (x'', y'') , of the input scene for which the RMSE is the smallest.
6. Extract a window centered on the pixel (x'', y'') , where the window size is given by an input parameter (256x256 in the experiments).
7. Return to step (2), if a predefined number of extracted chip-window pairs has not been met.

IV. ROBUST FEATURE MATCHING

Once a set of chip-window pairs has been automatically obtained, we employ a (hierarchical) robust feature matching (RFM) procedure that was presented in [11], [12]. The idea is to use a multiresolution overcomplete wavelet decomposition scheme to extract chip and window features at the various levels of the decomposition. Specifically, we perform a Simoncelli steerable decomposition of the chip and of the window (at each resolution level), and extract the pixels whose bandpass filter magnitudes are in the top 10%. Starting at the lowest (coarsest) resolution level, the RFM computes a match transformation that is iteratively refined at the higher levels.

Unlike typical applications of model-based pattern matching, which involve searching large spaces for small numbers of likely matches, image registration in the context of the above hierarchical scheme begins (at all levels) with *a priori* information about the expected bounds of the desired transformation. Such information is derived essentially from bounds on the uncertainty in the position or orientation of the imaging system at each level of decomposition. (See [12] for further details.)

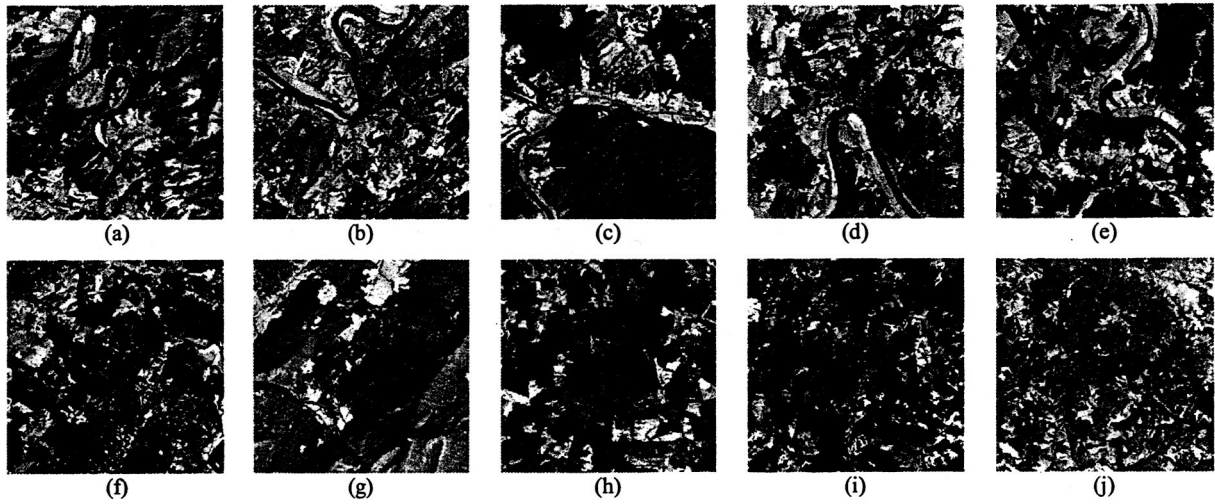


Figure 1. Ten chips extracted from a Landsat ETM+ scene collected over Central Virginia on October 7, 1999 (reference scene).

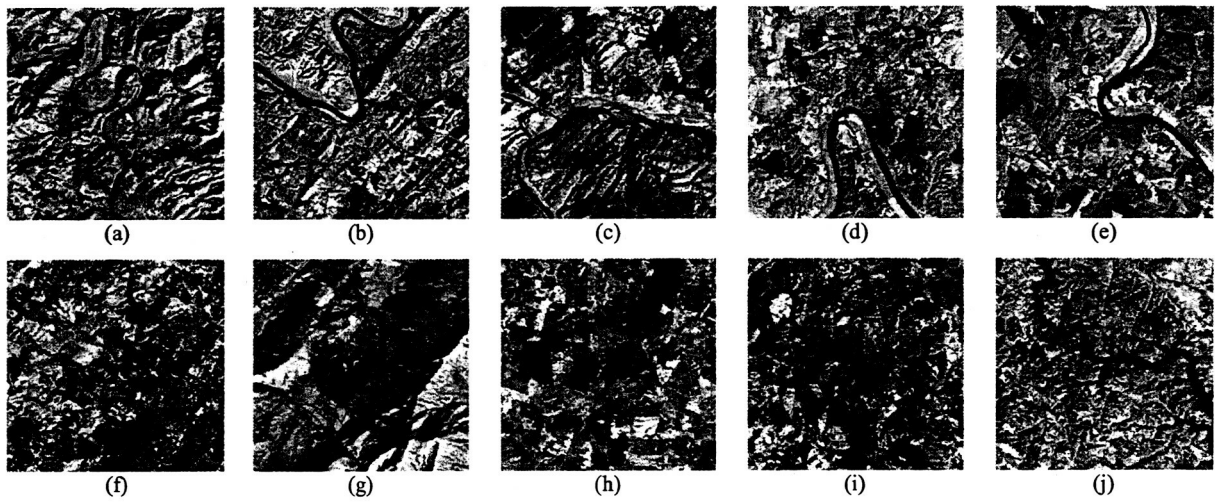


Figure 2. Ten windows extracted from a Landsat ETM+ scene collected over Central Virginia on November 8, 1999 (input scene).

TABLE I
REGISTRATION RESULTS OBTAINED FOR TEN CHIP-WINDOW PAIRS FROM TWO LANDSAT SCENES IN CENTRAL VIRGINIA

Chip-window	Rotation (deg)	Initial shift x	Initial shift y	Adjusted shift x	Adjusted shift y
(a)	0.0	12.0	4.0	14.0	1.0
(b)	0.0	-1.0	0.0	13.0	1.0
(c)	0.0	2.0	-2.0	13.0	1.0
(d)	0.0	0.0	0.0	13.0	1.0
(e)	0.0	-1.0	3.0	13.0	1.0
(f)	0.0	59.0	1.0	14.0	1.0
(g)	0.0	-3.0	-3.0	13.0	1.0
(h)	0.0	0.0	0.0	13.0	2.0
(i)	0.0	36.0	4.0	14.0	2.0
(j)	0.0	2.0	-2.0	13.0	1.0

For each chip-window pair, we perform robust matching of the selected wavelet-like features using the partial Hausdorff distance similarity measure and a computationally efficient search strategy based on the combination of geometric branch-and-bound and fast point-to-point alignments (see [11] for exact details). We use the matching to compute the parameters of a local transformation composed of a rotation and a translation. We then use its local transformation (for each pair) to determine the corrected locations of, say, the upper left corner of the window. The list of (upper-left) chip corners and of the corresponding corrected window corners serve as inputs to robust estimation of the parameters of a global transformation of the entire input scene. (See [12] for exact details concerning this robust postprocessing step.)

V. EXPERIMENTAL RESULTS

We have tested the fully automated system described in this paper on Landsat-7 ETM+ data acquired over Central Virginia on October 7, 1999. We report here on the results obtained for an incoming scene collected on November 8, 1999. Ten 256x256 chip-window pairs were extracted from the reference scene using the MM-based algorithm described in subsection III-B, using $k = 5$ scales and $l = 8$ orientations (The latter are additional input parameters.) The resulting reference chips are shown in Fig. 1(a)—(j), according to the order of their extraction by the algorithm, i.e., Fig. 1(a) shows the first chip that was produced, Fig. 1(b) shows the second chip, etc. Similarly, the corresponding windows that were extracted for these chips are shown in Fig. 2(a)—(j), respectively.

After applying RFM to the chip-window pairs in Figs. 1 and 2, we observed that, for the most part, the extracted chip-window pairs were initially fairly closely matched. For those pairs that were relatively far apart (on the order of tens of pixels), e.g. the pairs of Fig. 1(f)—Fig. 2(f), Fig. 1(g)—Fig. 2(g), and Fig. 1(i)—Fig. 2(i), the RFM still provided a good match. As a result, the ten local transformations obtained were very similar, suggesting a global transformation of no rotation and of 13- and 1-pixel translations in x and y , respectively. Indeed, the transformation due to the robust postprocessing step was (Rotation, Shift- x , Shift- y) = (0.0, 13.0, 1.0). The results above indicate that the proposed automatic registration approach can produce accurate results without prior knowledge. Further results and accuracy assessment (versus ground truth) will be presented at the workshop.

VI. CONCLUSIONS

In this paper, we have described a fully automated system for registration of multi-temporal Landsat data sets. The system can effectively deal with any size images (a typical size of a Landsat scene is on the order of 7500x8000 pixels) by registering a number of small windows extracted from the scene to their corresponding reference chips. Most importantly, the proposed system incorporates a

morphological-based algorithm for feature (and region-of-interest) extraction. The algorithm selects automatically a set of significant, evenly distributed chip-window pairs, that are centered on visually significant landmarks such as edges, corners and line intersections. Experimental results have demonstrated that the proposed morphological algorithm produces chip-window pairs that were initially fairly closely matched. Although some pairs were relatively far apart (on the order of tens of pixels), this initial approximation was successfully refined by employing a robust feature matching procedure followed by robust postprocessing to obtain the global transformation for the entire scene.

One of the most significant advantages of the proposed automatic system is that it eliminates the need of maintaining a chip database that would have to be updated regularly and adapted to each sensor. Also, the proposed morphological algorithm provides an initial condition which is relatively close to the final transformation, thus enabling optimization strategies to perform more accurately and with significant savings in computational time. Although experimental results obtained thus far are promising, further experiments and accuracy assessment, using additional multi-source data sets, are required in order to fully demonstrate the above remarks.

REFERENCES

- [1] L. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 326—376, 1992.
- [2] B. Zitova and J. Flusser, "Image Registration Methods: A Survey," *Image and Vision Computing*, vol. 21, pp. 977—1000, 2003.
- [3] J. Le Moigne, H. Stone, A. Cole-Rhodes, R. Eastman, P. Jain, K. Johnson, J. Morisette, N. S. Netanyahu, and I. Zavorin, "A Study of the Sensitivity of Automatic Image Registration Algorithms to Initial Conditions," *IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, September 20-24, 2004.
- [4] J. Serra, *Image Analysis and Mathematical Morphology*, Academic, New York, 1982.
- [5] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed., Springer, Berlin, 2003.
- [6] P. Soille and M. Pesaresi, "Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 2042—2055, 2002.
- [7] M. Pesaresi and J. A. Benediktsson, "A New Approach for the Morphological Segmentation of High Resolution Satellite Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 309—320, 2001.
- [8] A. Plaza, P. Martinez, R. Perez, J. Plaza, "Hyperspectral Image Analysis by Scale-Oriented Morphological Profiles," in *Proc. SPIE, Image and Signal Processing for Remote Sensing IX*, vol. 5238, pp. 432—439, 2003.
- [9] A. Plaza, P. Martinez, J. Plaza and R. Perez, "Dimensionality Reduction and Classification of Hyperspectral Image Data Using Sequences of Extended Morphological Transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 466—479, 2005.
- [10] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Kluwer, New York, 2003.
- [11] D. M. Mount, N. S. Netanyahu, and J. Le Moigne, "Improved Algorithms for Robust Point Pattern Matching and Applications to Image Registration," *Pattern Recognition*, vol. 32, pp. 17—38, 1999.
- [12] N. S. Netanyahu, J. Le Moigne, and J. G. Masek, "Georegistration of Landsat Data via Robust Matching of Multiresolution Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 1586—1600, 2004.