AN AREA BASED TECHNIQUE FOR IMAGE-TO-IMAGE REGISTRATION OF MULTI-MODAL REMOTE SENSING DATA

Peter Bunting, Richard Lucas
Institute of Geography and Earth Science
Aberystwyth University
Ceredigion, SY23 2DB, UK
pete.bunting@aber.ac.uk

Frédéric Labrosse
Department of Computer Science
Aberystwyth University
Ceredigion, SY23 2DB, UK

ABSTRACT

To allow remotely sensed datasets to be used for data fusion, either to gain additional insight into the scene or for change detection, reliable spatial referencing is required [1]. With modern remote sensing systems, reliable registration can be gained by applying an orbital model or through the use of a global positioning system (GPS) and inertial navigation system (INS) in the case of airborne and spaceborne datasets respectively. Whilst individually, these datasets appear well registered when compared to a second dataset from another source (e.g., Optical to LiDAR or Optical to SAR), the resulting images may still be several pixels out of alignment. Manual registration techniques are often slow and labour intensive and although an improvement in registration is gained, there can still be some misalignment of the datasets.

A large number of techniques for automating image-to-image registration have therefore been developed [2], with these generally implementing one of two approaches. The first (feature matching) aims to extract and match features across the image as two independent steps [3, 4] whilst the second (area matching) uses a metric to match regions of the image without explicitly extracting features [5, 6].

This study extended the area matching technique outlined in [7] by introducing the image similarity metrics based on a joint histogram. The original study used the correlation coefficient (\(cc\)) for identifying matching image regions but previous studies [2, 8] have demonstrated that joint histogram based techniques are potentially more reliable for matching multi-modal datasets. The algorithm developed by Bunting et al., [7] used a series of steps to produce the final registration, the first of which was to produce a pair of image pyramids from the original image data of matching image resolutions. Image matching started at the top of the image pyramid, the lowest resolution, and refined the registration as it descended the pyramid. A regular grid of tie points was defined in the overlapping region and at each level of the pyramid, where the overlapping region of the images was defined by the geographic metadata associated with each file. To control the movement of the tie points and maintain the topological relationships between the neighbouring tie points, a network structure was constructed between the tie points at each level of the pyramid and between the pyramid levels. Tie points were matched individually where a optimum was identified within a user-restricted search space formed by the similarity metric and through an exhaustive search. Following the identification of a new tie point position, the network was updated by propagating the identified transformation to neighbouring tie points using an inverted weighted distance.

The quantitative assessment of the accuracy of image registration has often proved to be a difficult task [2] because of the lack of ground truth data. This study adopted for the same approach and datasets as Bunting et al., [7] where two images pairs for each pair of selected datasets were manually registered and used for testing. The datasets used were 1 m LiDAR to 1 m CASI, 1 m LiDAR to 2.6 m HyMap, 1 m LiDAR to 3 m AIRSAR, 2.6 m HyMap to 1 m CASI, 2.6 m HyMap to 3 m AIRSAR and 25 m Landsat to 18 m ALOS-PALSAR. Four tests where carried out where the first used single modality image pairs without any introduced transformation. The other three used the multi-modal image pairs, outlined above, and the second test introduced transformation. The third introduced a series of translations in the \(X\) and \(Y\) axis’, from 2 pixels in the \(X\) axis and 3 pixels in the \(Y\) axis to 16 pixels in the \(X\) axis to 18 pixels in the \(Y\) axis results. The final test introduced non-linear sine wavelength independently into the \(X\) and \(Y\) axis’ with amplitudes of 5, 10 and 15. The results of these tests are given in Table 1, where the first row provides the results using \(cc\) from the previous study while the remaining rows provides the results for the joint histogram based measures distance to independence (\(d2i\)), Kolmgorov distance (\(kolm\)) and mutual information (\(mi\)). As the transformation between each image pair was known, the accuracy of the registration could be calculated by generating a set of correct tie points for each experiment to which the tie points generated by the registration algorithm could be compared.
The comparison was made at the highest pixel resolution of the two images where the linear distance between the produced tie point to its correct location was calculated and the mean and standard deviation values provided across all tests.

Table 1. The results for automatic image-to-image registration using correlation coefficient, distance to independence, Kolmogorov distance and mutual information, where the mean and standard deviation (in pixels) are provided.

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>Single Modality</th>
<th>Multi-modal</th>
<th>Translation</th>
<th>Non-linear Sine</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc</td>
<td>0.011 (0.024)</td>
<td>3.465 (2.477)</td>
<td>4.980 (3.920)</td>
<td>7.616 (6.153)</td>
</tr>
<tr>
<td>d2i</td>
<td>0.006 (0.017)</td>
<td>12.996 (7.038)</td>
<td>50.840 (42.622)</td>
<td>54.402 (46.844)</td>
</tr>
<tr>
<td>kolm</td>
<td>0.0 (0.0)</td>
<td>19.194 (16.237)</td>
<td>54.429 (37.735)</td>
<td>65.590 (36.611)</td>
</tr>
<tr>
<td>mi</td>
<td>0.007 (0.017)</td>
<td>25.180 (14.0138)</td>
<td>66.238 (44.216)</td>
<td>67.674 (39.422)</td>
</tr>
</tbody>
</table>

These results indicated that, under these condition, the correlation coefficient was the more reliable similarity measure. Obtaining poorer results for the joint histogram measures was unexpected given that previous studies have suggested that the correlation coefficient should not work for multi-modal registration [2, 8]. In fact, Inglada and Giros [8] even suggested the joint histogram methods, and in particular mutual information, should be more robust than the correlation coefficient under all circumstances. In this study, it is believed that this discrepancy occurs because of the size of the error when an incorrect result is returned and the effect of the network used to control the tie points. If a large error is inputted to the network at the top level, this error will be propagated down through the network and between levels multiplied by the scale factor (between the levels within the pyramid). Therefore, any large errors entered at the top of the network can significantly move lower levels of the network out of alignment, generating large errors in the final registration. It is, therefore, considered that the joint histogram based measures (e.g., mi) create a small number of high magnitude errors when compared to the correlation coefficient and that these errors, particularly if they occur higher up in the network, are magnified by the network structure. Furthermore, as a regular grid of tie points is required rather than identifying areas of high similarity, in which tie points are identified, large errors are less avoidable than in other studies [8].

1. REFERENCES


