THE USE OF SIMILARITY IMAGES ON MULTI-SENSOR AUTOMATIC IMAGE REGISTRATION

H. Gonçalves [a,b] *, J. A. Gonçalves [a], L. Corte-Real [c,d]

[a] Departamento de Geociências, Ambiente e Ordenamento do Território, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n 4169-007 Porto, Portugal - (hernani.goncalves,jagoncal)@fc.up.pt
[b] Centro de Investigação em Ciências Geo-Espaciais, Universidade do Porto, Rua do Campo Alegre, s/n 4169-007 Porto, Portugal
[c] Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto, Rua Dr Roberto Frias s/n 4200-465 Porto, Portugal
[d] INESC Porto, Rua Dr Roberto Frias s/n 4200-465 Porto, Portugal – lreal@inescporto.pt

Commission VII

KEY WORDS: Automation, Correction, Correlation, Georeferencing, Image, Matching, Mathematics

ABSTRACT:

Automatic image registration (AIR) is still a present challenge regarding remote sensing applications. Although several methods have been proposed in the last few years, geometric correction is often a time and effort consuming manual task. The only AIR method which is commonly used is the correlation-based template matching method. It usually consists on considering a window from one image and passing it throughout the other, looking for a maximum of correlation, which may be associated to the displacement between the two images. This approach leads sometimes (for example with multi-sensor image registration) to low correlation coefficient values, which do not give sufficient confidence to associate the peak of correlation to the correct displacement between the images. Furthermore, the peak of correlation is several times too flat or ambiguous, since more than one local peak may occur. Recently, we have tested a new approach, which shortly consists on the identification of a brighter diagonal on a "similarity image". The displacement of this brighter diagonal to the main diagonal corresponds to the displacement in each axis. In this work, we explored the potential of using the "similarity images" instead of the classical "similarity surface", considering both correlation coefficient and mutual information measures. Our experiments were performed on some multi-sensor pairs of images with medium (Landsat and ASTER) and high (IKONOS, ALOS-PRISM and orthophotos) spatial resolution, where a subpixel accuracy was mostly obtained. It was also shown that the application of a low-pass filtering prior to the similarity measures computation, allows for a significant increase of the similarity measures, reinforcing the strength of this methodology in multi-spectral, multi-sensor and multi-temporal situations.

1. INTRODUCTION

Multi-sensor automatic image registration (AIR) is a present challenge, with emphasis on remote sensing applications. Direct georeferencing techniques, based on navigation instruments on board the satellites allow for the determination of pixel geolocation. Bringing images to a well defined cartographic reference system allows for an approximate image registration with any other precisely georeferenced imagery. Since ideally image registration should be done at least at the pixel accuracy, improvement is needed for most satellite images.

A wide variety of AIR methods may be found in the literature (Brown, 1992; Fonseca, 1996; Zitová, 2003). However, there are several particularities on the registration of remote sensing images which justifies continuous research in this field. These particularities include differences in the radiometric content (motivated by different spectral bands and/or different sensors), the slope variation of the terrain covered by the image, differences in the image acquisition geometry, among other difficulties. A system which should automatically analyse all these aspects and select the most appropriate method or a combination of methods seems to be the most reasonable solution for the complex problem of multi-sensor AIR.

The most popular methods for AIR are those based on similarity measures, where the correlation coefficient plays an important role (Inglada and Giros, 2004). This class of methods mainly consists on taking a template from an image and pass it throughout the other image, producing a similarity surface. The shift between the images is expected to be associated to a well defined peak on the similarity surface. However, in several times, the surface peak may be associated to a low correlation value, present a smooth peak leading to a less accurate location, or even erroneous peaks may be found.

For the above mentioned facts, (Gonçalves et al., 2008) proposed an automatic image registration method based on correlation and Hough transform, which allows for reducing these weaknesses associated to the traditional approach of correlation-based methods. In this work, this approach was further explored by also considering the mutual information, as well as an analysis regarding the computational time, evaluated for different pairs of images. The proposed methodology is described in section 2, and some examples of its application are provided in section 3. The discussion and conclusions correspond to sections 4 and 5, respectively.

* Corresponding author.

2. METHODOLOGY

A new approach for the use of the correlation coefficient in automatic image registration was recently explored (Gonçalves etal., 2008). In this paper, we generalized this approach, which will be described in the following. In order to simplify the provided analysis, we will focus on the problem of finding a translation in both horizontal and vertical directions, assuming that the considered region is approximately "flat". Considering (P_{REF} ,L_{REF}) and (P_{NEW},L_{NEW}) as the (Pixel,Line) coordinates of the reference and new (to be registered) images, respectively, their relation may be expressed as

$$P_{NEW} = P_{REF} + \delta_x \quad (1)$$

$$L_{NEW} = L_{REF} + \delta_y \quad (2)$$

where δ_x and δ_y are the displacements (in pixel units) on the horizontal and vertical directions, respectively, between the reference and the new image. The registration of a full scene or images with more complex deformations may be performed according to the description in (Gonçalves etal, 2008), and further evaluated trough a proper set of measures (Gonçalves etal., 2009). The several steps of the proposed methodology will be described in the following.

2.1 Division of the image into tiles

As previously mentioned, in this work the focus relies on approximately "flat" regions. Depending on the terrain slope variation and on the image acquisition geometry, it may become difficult to avoid slight differences on the shifts throughout the images. Therefore, the division of the image into tiles is also considered in this work, to evaluate whether it may allow for reducing some of these remaining effects. The tiles must be sufficiently higher than the shift known or estimated a priori, for which a minimum size of 64x64 pixels up to the full image size (a single tile) may be generally applicable. The following steps are applied to each tile.

2.2 Similarity image

Instead of the traditional similarity surface, two similarity images are proposed, each one corresponding to the horizontal and vertical directions. The similarity image is produced by computing a similarity measure along one dimension at a time. Considering a tile with m-by-n pixels, then the similarity image for the horizontal and vertical directions will have n-by-n and m-by-m pixels, respectively. This procedure is schematically represented in Figure 1. The correct shift between the tiles is expected to produce a brighter diagonal strip on the similarity image, corresponding to the higher values of the similarity measure. An example of a similarity image is provided in Figure 2c.

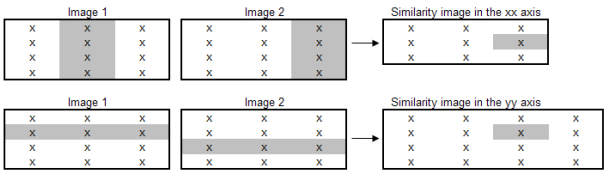

Figure 1. Illustration of the similarity image computation in both xx and yy axis.

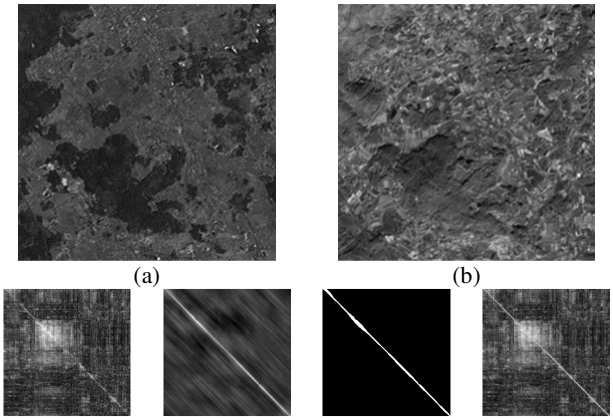

Figure 2. Illustration of the main steps of the proposed methodology (further details in section 2): (a) A segment with 256×256 pixels from a Landsat image; (b) a segment with 256×256 pixels from an ASTER image; (c) similarity image (using the correlation coefficient); (d) filtered similarity image; (e) image in (d) converted to binary; (f) –45º line detected by the Hough transform, superimposed on the similarity image.

2.3 Similarity image filtering

In order to enhance the visibility of the similarity image brighter diagonal strip, the similarity image is filtered using a –45º oriented line. The length of this line is defined to be 15% of the image dimension norm, which may be broadly applied to any sensor (Gonçalves etal., 2008). The filter window is composed by positive values along the diagonal and zeros outside. Its effect is illustrated in Figure 2d.

2.4 Conversion from gray level to binary

Prior to the Hough transform computation, there is the need to convert the filtered similarity image to a binary format. A threshold equal to the percentile (1–3/n)×100 (rounded to the smaller integer) is considered, where n is the number of lines (or columns) of the similarity image (which is squared). The binary image of the example is provided in Figure 2e.

2.5 Hough transform

At the Hough transform step (Hough, 1962), the θ and ρ resolution is defined as 0.5. The Hough transform is computed for the similarity images in both xx and yy axis. For each of them, more than one line may be identified, associated to the most prominent peaks.

2.6 Main diagonal identification and displacement computation

The slope of the detected line(s) in the previous step is computed, being considered only those with slope between -0.95 and -1.05. In case of more than one line is detected with a slope of exactly -1.00, the line with highest height is selected. The displacement on each axis is finally obtained by computing the distance from the selected line to the main diagonal (–45º line starting at row 1 column 1). This step is illustrated in Figure 2f.

2.7 Estimation of δ_x and δ_y

In the case that the image is not divided into tiles (the image being itself a single tile), then the estimates for δ_x and δ_y are

merely the displacements obtained at the previous described step. When more than one tile is considered, then a set of candidates for δ_x and δ_y are obtained, which inevitably may contain some misleading values. Therefore, an outlier removal stage is required, in order to consider only those correct shifts and achieve acceptable (subpixel) accuracy. The estimation of δ_x and δ_y consisted on a statistical procedure based on the 3D histogram obtained from δ_x and δ_y.

3. RESULTS

3.1 Dataset

The proposed methodology was applied to three pairs of images, comprising medium and high (urban and urban/rural context) spatial resolution images (Figure 3). The digital elevation model (DEM) of the considered regions obtained from the shuttle radar topography mission (SRTM) (Farr, 2004) is presented in Figure 4.

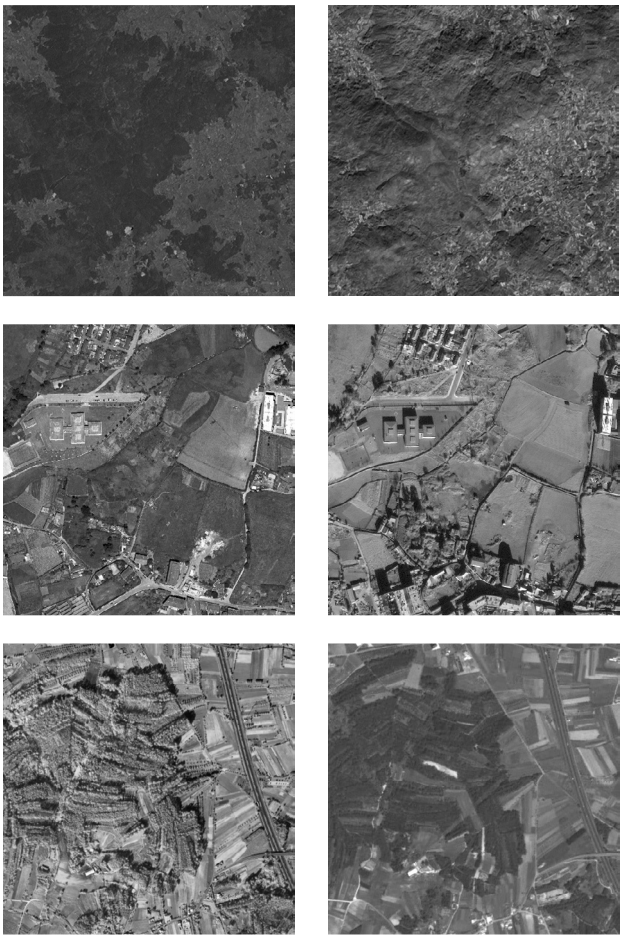

Figure 3. Reference images (left) and new images to be registered (right), regarding a medium spatial resolution pair (first row, Landsat/ASTER), a high spatial resolution pair with urban/rural context (second row, Orthophoto/IKONOS) and a high spatial resolution pair with rural context (third row, Orthophoto/ALOS). Further details in subsection 3.1.

3.1.1 Medium spatial resolution

The first pair of images consists in two segments with 512x512 pixels: one obtained from an orthorectified panchromatic Landsat image; and the other from an ASTER image (NIR band) with an approximate geometric correction. Both of these images are from the northwest of Portugal. Both images have a pixel size of 14.25m and present a temporal difference of 1.5 years. The two segments are represented in Figure 3 (first row).

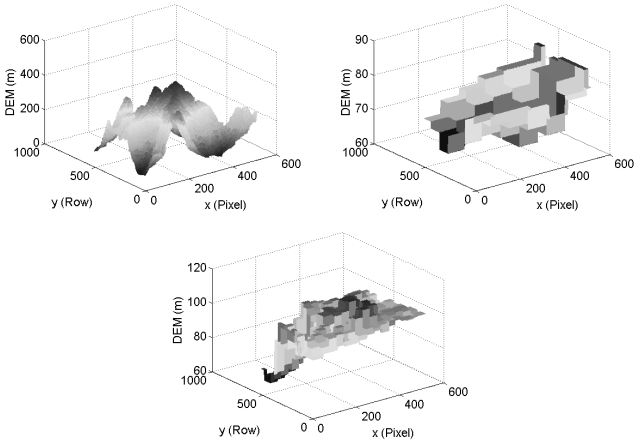

Figure 4. DEM (based on the SRTM) associated to the Landsat/ASTER pair (upper left), orthophoto/IKONOS pair (upper right) and Orthophoto/ALOS pair (bottom). Further details in subsection 3.1.

As can be observed in Figure 4, the considered region for this pair of images presents a considerable terrain height variation, ranging from 29m to 487m (according to the SRTM DEM). However, the transformation function defined in (1) and (2) may be considered adequate due to the ASTER acquisition geometry. The reference shifts were manually obtained through the identification of 4 conjugate points, associated to an average±standard-deviation of 2.1±0.1 and 1.4±1.3 pixels, for the horizontal and vertical directions, respectively. The low standard-deviation of the manually identified conjugate points support the adequacy of the transformation function.

3.1.2 High spatial resolution (IKONOS): urban/rural context

The second pair of images is composed by two segments with 512x512 pixels: one corresponding to the green band of an orthophoto; and the other corresponding to a panchromatic IKONOS image with an approximate geometric correction. These images cover a small part of the city of Porto (Portugal) with a mixture of urban with rural context, have a pixel size of 1m and present a temporal difference of around 1 year. The two segments are represented in Figure 3 (second row).

As can be observed in Figure 4, the considered region for this pair of images presents a smooth terrain elevation, ranging from 67m to 87m (according to the SRTM DEM). The reference shifts were manually obtained through the identification of 4 conjugate points, associated to an average±standard-deviation of 15.9±1.8 and 5.7±0.2 pixels, for the horizontal and vertical directions, respectively.

3.1.3 High spatial resolution (ALOS): rural context

The third pair of images comprises two segments with 512x512 pixels: one corresponding to the NIR band of an orthophoto; and the other corresponding to a panchromatic ALOS-PRISM image with an approximate geometric correction. These images are from the centre of Portugal (rural context), have a pixel size

of 2.5m and present a temporal difference of around 2 years. The two segments are represented in Figure 3 (third row).

As can be observed in Figure 4, the considered region for this pair of images presents a smooth terrain elevation, ranging from 68m to 114m (according to the SRTM DEM). The reference shifts were manually obtained through the identification of 5 conjugate points, associated to an average±standard-deviation of -4.6±0.9 and 0.1±0.8 pixels, for the horizontal and vertical directions, respectively.

3.2 The traditional approach

The traditional approach of AIR based on similarity measures mainly consists on taking a window (template) from one image and pass it throughout the other, aiming to find a peak on the similarity surface. This peak is expected to correspond to the correct shift (in both horizontal and vertical directions) between the images. The location of the template, the size of the template, and the associated computational cost may lead to a wide variety of template selections. The results presented in Figure 5 illustrate the effect of considering different sizes of the template (for the three pairs of images in Figure 3) - defined as a squared region with its centre corresponding to the center of the reference image – considering the correlation coefficient as the similarity measure. Although the computational time increases with the increase in the template size, it still presents a relatively fast performance (Figure 6). The ambiguous aspect associated to the template selection may lead to misleading solutions, as the results presented in Figure 5 clearly illustrate, in particular for the medium spatial resolution images. Furthermore, for the high resolution images, the traditional approach is not able to accurately register them.

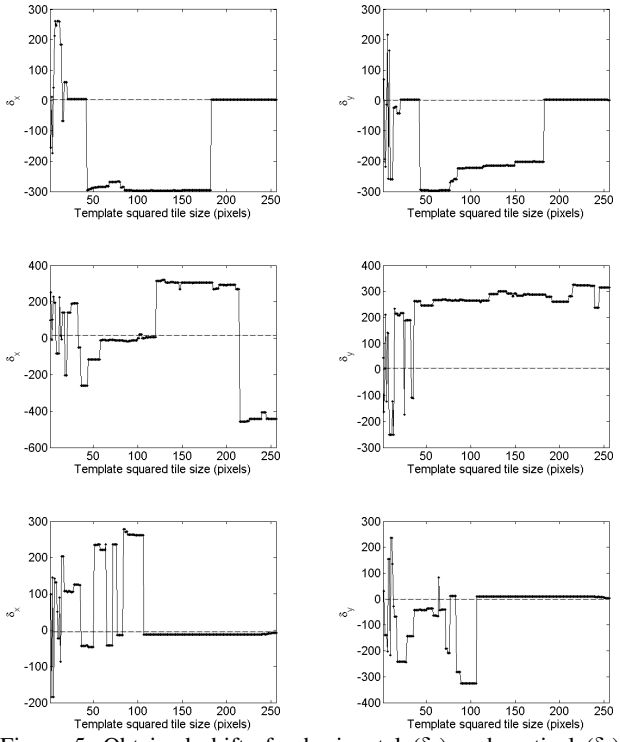

Figure 5. Obtained shifts for horizontal (δ_x) and vertical (δ_y) directions on the first and second columns, respectively, using the traditional approach described in subsection 3.2, applied to the three pairs of images presented in Figure 3 (in the same order from top to bottom). Dashed lines are the reference shifts.

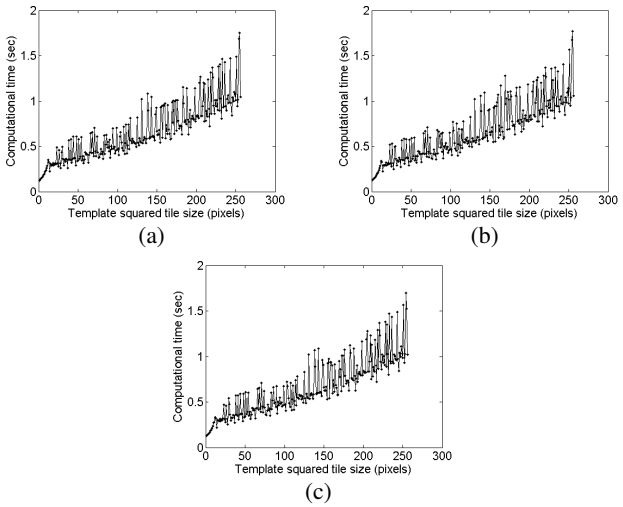

Figure 6. Computational time (in seconds) associated to the traditional approach as described in subsection 3.2, applied to the three pairs of images presented in Figure 3: (a) Landsat/ASTER; (b) orthophoto/IKONOS; (c) orthophoto/ALOS.

3.3 Application of the proposed methodology

There is a wide variety of similarity measures which may be applied in the proposed methodology (Inglada and Giros, 2004). The correlation coefficient (CC) is one of the most used similarity measures regarding image registration applications, and its definition is widely known (Brown, 1992; Inglada and Giros, 2004; Zitová and Flusser, 2003). The mutual information (MI) of two random variables A and B can be obtained as (Cover, 1991)

$$MI(A,B) = H(A) + H(B) - H(A,B) \quad (3)$$

where H(A) and H(B) are the entropies of A and B, and H(A,B) is their joint entropy. The MI-based registration criterion states that the images shall be registered when MI(A,B) is maximal. The remaining definitions of the entropies and corresponding probabilities can be found in (Chen, 2003).

The CC and MI measures were applied to the pair of images represented in Figure 2, considering each image as a single tile. The obtained similarity images for both horizontal and vertical directions are provided in Figure 7. It can be observed that the CC is clearly more adequate than MI. One of the reasons behind this may be the fact that we have applied cross-correlation to all possible lags, and used the maximum among these. This procedure allows for minimizing the misalignment which is present when computing 1D correlation. For instance, when computing the correlation on the horizontal direction, the DNs values of each column from the reference image will present some misalignment on the corresponding column of the image to be registered, due to the shift on the vertical direction. Additionally, the CC presents a significant faster performance than MI.

Based on the above mentioned experiments, in this work the application of the proposed methodology will rely on the CC as the similarity measure. The obtained results for both horizontal (δ_x) and vertical (δ_y) directions, with respect to the three pairs of images in Figure 3 are provided in Figure 8 (considering tiles of size 64x64, 128x128, 256x256 and 512x512 pixels).

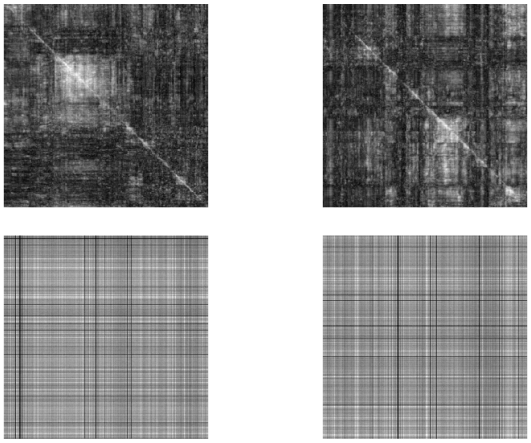

Figure 7. Similarity images (horizontal and vertical directions on the first and second column, respectively), regarding the registration of the pair of images represented in Figure 2, considering as similarity measure the CC (first row) and the MI (second row).

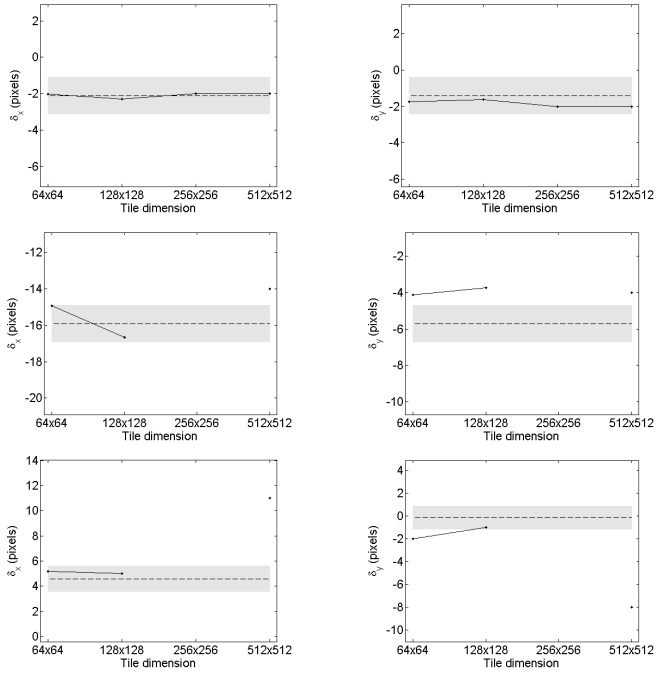

Figure 8. Accuracy in the horizontal (δ_x, left plots) and vertical (δ_y, right plots) directions regarding the registration of the three pairs of images represented in Figure 3 (a single tile), using the CC as the similarity measure: Landsat/ASTER (first row); Orthophoto/IKONOS (second row); Orthophoto/ALOS (third row).

Regarding the medium spatial resolution pair of images, the proposed methodology was able to achieve a subpixel accuracy for all considered tile sizes (Figure 8). The traditional approach was able to achieve a similar accuracy for templates higher than 190x190 pixels, leading to misleading results for smaller templates (Figure 5). Therefore, the proposed methodology presents clear advantages with respect to the traditional approach.

With respect to the high spatial resolution pair of images with urban/rural context (orthophoto/IKONOS), a subpixel accuracy was obtained for tiles with size 64x64 and 128x128 pixels in the horizontal axis, and an error less than 2 pixels in the vertical axis. Despite the error above the pixel in the vertical direction, it is quite obvious the advantage when compared to the traditional approach, which was quite far from the correct solution for all possible template sizes. For tiles with size 256x256 pixels, the proposed methodology did not provide any solution, which is better than indicating a wrong solution. Even considering the whole images as a single tile, the obtained shifts were quite near the reference values. Moreover, even the manual identification of conjugate points was associated to a standard deviation of 1.8 pixels on the horizontal direction, supporting the difficulty of accurately registering this pair of images.

For the third pair of images, composed by two high spatial resolution segments with rural context (orthophoto/ALOS-PRISM), a subpixel accuracy was also obtained for tiles with size 64x64 and 128x128 pixels. Once again, for tiles with size 256x256 pixels no solution was obtained, which is better than a wrong solution. Considering a single tile (512x512 pixels), an error of around 6 pixels was obtained for both directions. This result indicates that a single tile should be avoided, since when using smaller tiles the statistical based procedure of outliers removal provides generally an accurate registration. Nevertheless, although the traditional approach tends to achieve an accurate solution for templates with size near the whole image, a closer look at the plots in Figure 5 allows for observing that the best obtained results are 4 pixels far from the correct solution. Therefore, the proposed methodology is once again generally better than the traditional approach.

With respect to the computational efficiency (Figure 9), it can be observed that beyond the smaller tiles provide more accurate results, they are also associated to lower computational times, reinforcing their advantage. Although the presented computational times are considerably higher than the traditional approach (Figure 6), it is worth to mention that it was not under the scope of this work the computational time optimization. However, it can be largely improved, since several graphical outputs which are produced and stored are unnecessary to provide the final estimates of δ_x and δ_y.

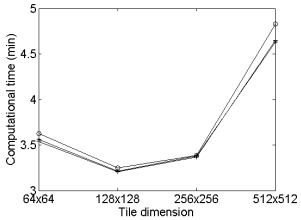

Figure 9. Computational time associated to the results obtained in the registration of the three pairs of images represented in Figure 3 (for different tiles dimension), using the CC as the similarity measure: Landsat/ASTER (+); orthophoto/IKONOS (*); orthophoto/ALOS (o).

3.4 Future improvements

The proposed approach revealed to outperform the traditional approach of image registration using similarity measures, in particular for images with clearly different radiometric content. Nevertheless, some preliminary tests have been performed which may allow for further improvements in the future. A visual inspection from the upper plots in Figure 10 suggest that a low-pass filtering may allow for extracting a profile with less detail, ignoring higher variability related to the spectral characteristics. Therefore, a 2nd–order low-pass Butterworth filter, with a normalized cutoff frequency at 0.1 was applied to

the original profiles, which result is provided in the lower plots of Figure 10. It can be observed that a noticeable improvement on the similarity between the profiles from the Landsat and ASTER images was obtained, supported by a considerable increase of both similarity measures. The potential of this filtering stage deserves further research, in particular with respect to the increase in the computational complexity and subsequent processing time.

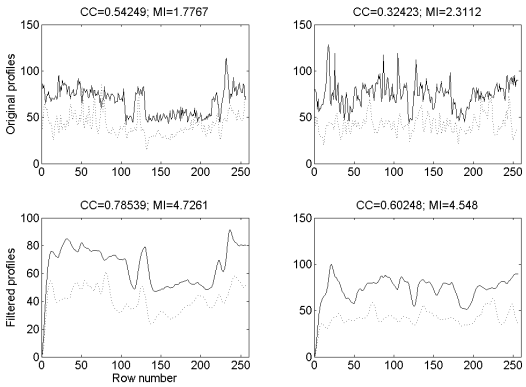

Figure 10. Two different profiles (left and right plots) obtained from the images in Figure 2: solid lines are columns 100 and 250 from the Landsat image; dotted lines are columns 101 and 251 from the ASTER image. The plots from the first row correspond to the original profiles, whereas the lower plots are the same profiles after low-pass filtering. The CC and MI similarity measures are provided above each plot. Further details in subsection 3.4.

4. DISCUSSION

The proposed methodology starts with the division of an image into tiles. With respect to the tiles dimension, it was observed that the smaller tiles led in general to more accurate results. This is related to the fact that when using smaller tiles, a larger set of shifts are obtained. Although a higher number of misleading shifts may be obtained, the statistical based procedure of outliers removal allow for focusing on a "cloud" of correct shifts.

It was shown that the proposed methodology clearly outperforms the traditional approach of using similarity measures on image registration. It should be noticed that through the division of the image into tiles, it was possible to achieve a subpixel accuracy, without requiring the use of fractional shifts.

Although accurate results were obtained using the CC, other similarity measures than the CC and MI could have been used and will deserve further research. In particular, the cross-ApEn (Pincus and Singer, 1996) which is an entropy-based measure will be explored in the future.

The proposed methodology comprises some image processing steps, which are necessarily associated to higher computational costs. However, the presented computational times are far from being optimized, since a large number of graphical outputs which are produced and stored for quality assessment are totally unnecessary for what really cares, which is merely the estimation of δ_x and δ_y. Therefore, further work on this topic will allow for a drastic reduction of the presented computational times.

The selection of the dataset segments was based on considering regions with slight terrain slope variations. However, there were still some significant slopes, which may lead to less accurate results. In the case that smaller tiles are used, the set of obtained shifts may be weighted by a cost function associated to the terrain height variation in each tile. This is an idea which deserves further research.

5. CONCLUSIONS

A new approach for the use of similarity measures was explored, which allows for an accurate registration of multi-sensor, multi-spectral and multi-temporal pairs of remote sensing images. It allows for reducing the ambiguity associated to the traditional approach, providing robust estimations of both horizontal and vertical shifts. A set of local shifts may be used for the registration of full scenes with more complex distortions.

ACKNOWLEDGEMENTS

The first author acknowledges Fundação para a Ciência e a Tecnologia, Portugal, for the financial support.

REFERENCES

Brown, L.G., 1992. A survey of image registration techniques. *Computing surveys* 24(4), pp. 325–376.

Chen, H., Varshney, P.K., Arora M.K., 2003 Performance of mutual information similarity measure for registration of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 41(11), pp. 2445-2454.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. New York: Wiley.

Farr, T.G., et al., 1992. The shuttle radar topography mission. *Reviews of Geophysics* 45, pp. RG2004.

Fonseca, L.M.G., Manjunath, B.S., 1996. Registration techniques for multisensor remotely sensed imagery. *Photogrammetric Engineering & Remote Sensing* 62(9), pp. 1049–1056.

Gonçalves, H., Gonçalves, J.A., Corte-Real, L., 2008. Automatic image registration based on correlation and Hough transform. In: *Image and Signal Processing for Remote Sensing XIV*, Cardiff, Wales, Vol. 7109. pp. 71090J.

Gonçalves, H., Gonçalves, J.A., Corte-Real, L., 2009. Measures for an objective evaluation of the geometric correction process quality. *IEEE Geoscience and Remote Sensing Letters*, 6(2), pp. 292–296.

Hough, P.V.C., 1962. Methods and means for recognizing complex patterns. U.S. Patent 3.069.654, 1962.

Inglada, J., Giros, A., 2004. On the possibility of automatic multisensor image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10), pp. 2104–2120.

Pincus, S., Singer, B.H., 1996. Randomness and degrees of irregularity. *Proc. Natl. Acad. Sci. USA*, 93(3), pp. 2083–2088.

Zitová, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing*, 21(11), pp. 977–1000.