ONE-TO-MANY REGISTRATION OF LANDSAT IMAGERY

L. Barazzetti [a], M. Scaioni [b,*], M. Gianinetto [a]

[a] Politecnico di Milano, Department of Architecture, Built Environment and Construction Engineering
Via Ponzio 31, 20133 Milano, Italy
(luigi.barazzetti, marco.gianinetto)@polimi.it
[b] Tongji University, College of Surveying and Geo-Informatics, Center for Spatial Information Science and Sustainable
Development Applications, 1239 Siping Road, 200092 Shanghai, P.R. China
marco@tongji.edu.cn - * corresponding author

Commission VII-6

KEY WORDS: Automation, Landsat, Matching, Registration, Satellite imagery

ABSTRACT:

This paper presents an innovative methodology for the registration of multiple satellite scenes by using a global Least Squares adjustment. The main advantage in relation to traditional approaches concerns the combined use of all the images of the dataset to obtain a more robust and comprehensive registration along with a statistical evaluation of the solution accuracy. This technique avoids standard matching solutions (implemented in several commercial software packages) where pairs of images are independently co-registered on the basis of an 'one-to-one' approach, starting from points extracted via independent 'image-to-master' matching. Once a set of corresponding multi-image features is extracted from the whole dataset, the implemented algorithm provides a global mapping function for an adjustment process that simultaneously includes all the available data. The multi-image matching process (coined 'one-to-many' approach) is performed by exploiting multi-temporal image combinations to obtain features visible in as many scenes as possible. The features are then clustered to generate a regular structure of image coordinates for the following Least Squares adjustment phase. The method is able to manage features visible in as many images as possible and is therefore a powerful alternative for registering satellite data which do not directly share common features with the master. Some examples will be illustrated to report the feasibility of the registration algorithm and to prove its sub-pixel accuracy for the specific case of Landsat images.

1. INTRODUCTION

Landsat imagery are a powerful source of information for Earth Observation. They provide time series over a span of 40 years and remain the only data available for such a long period with a ground sampling distance (GSD) of some tens of meters, i.e. ranging from 80 m grid cell of the Multi-Spectral Scanner (MSS) up to 15 m spatial resolution for the panchromatic band of the Enhanced Thematic Mapper Plus (ETM+). In addition, new data are now available after the successful launch of Landsat 8 on February 11, 2013.

Nowadays, a huge amount of orthorectified Landsat images are available (free of charge) and can be used for remote-sensing applications thanks to the U.S. Geological Survey's Earth Resources Observation and Science Center (http://eros.usgs.gov) and NASA's Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov). Data are geo-referenced by using the UTM projection (ellipsoid WGS84) and are blocked into 5 degrees N-S partitions (Tucker et al., 2004).

Their geometric registration is an extremely important topic. Unregistered data cannot be used for Remote Sensing as image pixels of different objects do not represent the same object (Goshtasby 2005; Le Moigne et al., 2011; Gianinetto et al., 2012;). It is well known that small misalignments in the input data could lead to large errors in the final outputs (Townshend et al., 1992).

The registration of satellite images (image-to-image registration) is usually carried out with a set of control points (CPs), whose measurement is accomplished by following a quite standard workflow: (i) CP extraction from both 'reference'

(or 'source', or 'master') and 'sensed' (or 'target' or 'slave') images; (ii) estimation of the parameters of a selected mathematical model that maps the images; (iii) resampling of the sensed image. This approach is currently implemented in several commercial software packages through specific modules, e.g. MicroImages Auto-register , ENVI registration tool, Erdas AutoSynch, or PCI Geomatics. These software can work on specific bands or band combinations and use different mathematical models generally based on similarity, affine, homographic, or more complex geometric transformations (e.g., Rational Polynomial Coefficients - Poli and Toutin T., 2012).

The extraction of the CPs can be carried out *manually* or in a fully *automated* way. In the first case, an operator interactively selects a set of corresponding features with a good distribution. For the second solution (automated) different algorithms are available and mainly include correlation-like methods (Pratt, 1991), mutual information (Pluim et al., 2001), Fourier methods (Castro and Morandi, 1987), or relaxation methods (Price, 1985).

On the other hand, these methods (manual or automated) are based on the analysis of a single image pair. In the case of multiple images, a separated matching-registration process is carried out between a single reference and multiple sensed images. This matching strategy could be defined as an 'one-to-one' approach (Gianinetto and Scaioni, 2008) where features between 'sensed-to-sensed' image combinations (in the case of datasets made up of 3 or more images) are not taken into consideration.

This paper describes the first results of a new automatic multi-image registration technique where all the images of a time

series are used in a global adjustment, therefore improving the 'one-to-one' approach. This method provides a global mapping function within an adjustment process that simultaneously includes all the images of the data set in a 'one-to-many' manner (see the flowchart in Fig. 1). The image matching process can be performed manually (interactive extraction of multi-image features) or in an automated way with a matching strategy that exploits all the image pairs by using the SURF operator (Bay et al., 2008). The features are then clustered and reordered to generate a regular structure of image coordinates for the following global Least Squares (LS) adjustment. The implemented algorithm extracts points often visible in more than two images and the LS adjustment provides the final simultaneous estimation of all parameters. One image is selected as 'master' in order to define the datum where the remaining images will be registered. Consequently, the design matrix of the solving system of equations will include both 'master-to-slave' and 'slave-to-slave' equations, enabling the registration of images without a direct connection to the master (see next sections for more details).

A connection graph showing multiple links among the images is used to initialize a registration algorithm that maps every slave image onto the master of the project.

In this study a 2D *similarity transformation* (encompassing four parameters) was used as mapping model for the Landsat imagery, although the method can be extended towards more complex transformations to handle more complex data.
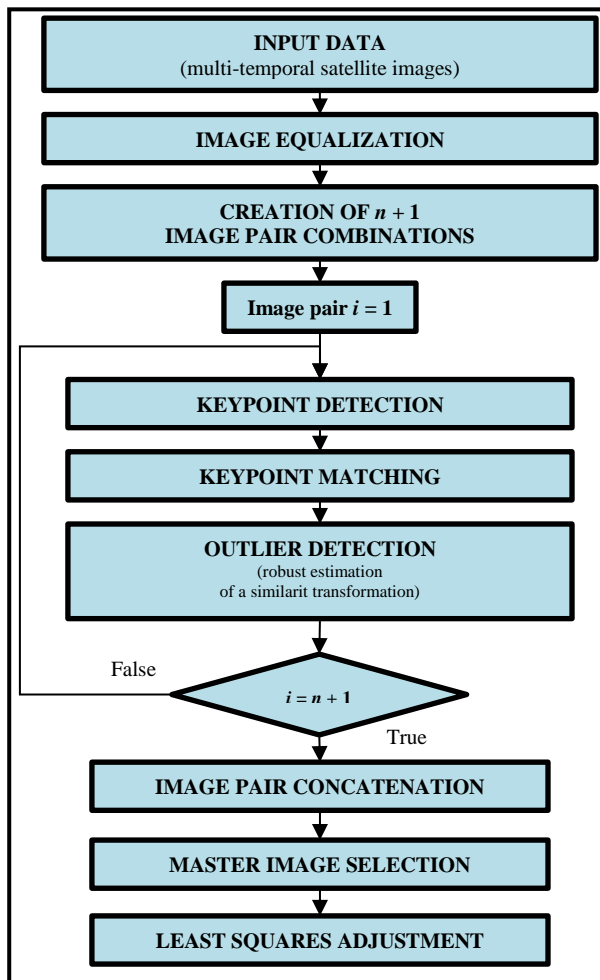

Figure 1. Flowchart of the algorithm for 'one-to-many' image registration.

2. SIMULTANOUS LEAST SQUARES REGISTRATION

As the method aims at registering multiple images by means of a global adjustment, corresponding points are needed not only for 'master-to-slave' image combinations, but also for all 'slave-to-slave' image pairs. These features can be interactively extracted through manual measurements or by using an automatic matching algorithm (see next section for further details).

For all the points matched on more than two images, a geometric registration technique that simultaneously estimates all parameters was developed. The proposed algorithm replicates the principle of photogrammetric bundle adjustment (Kraus, 2007), making the estimation more accurate and reliable, especially for time series incorporating several images. The geometric model used in this study is a similarity transformation relating the image point coordinates on the pair of images (x, y and x', y'):

$$x'_{ij} = s_j cos\alpha_j x_{ij} - s_j sin\alpha_j y_{ij} + c_j$$
$$y'_{iM} = s_j sin\alpha_j x_{ij} + s_j cos\alpha_j y_{ij} + d_j$$
(1)

where i is the point index and j the image index. After the substitution $a_j = s_j cos\alpha_j$ and $b = s_j sin\alpha_j$ two kind of equations can be written. The first one (Eq. 2) entails the transformation between a 'master-to-slave' pair:

$$x'_{iM} = \underline{a}_j x_{ij} - \underline{b}_j y_{ij} + c_j$$
$$y'_{iM} = \underline{b}_j x_{ij} + \underline{a}_j y_{ij} + d_j$$
(2)

The second one (Eq. 3) concerns 'slave-to-slave' matches:

$$\underline{a}_j x_{ij} - \underline{b}_j y_{ij} + c_j - \underline{x}'_{jj} = 0$$
$$\underline{b}_j x_{ij} + \underline{a}_j y_{ij} + d_j - \underline{y}'_{ij} = 0$$
(3)

where the underlined quantities are the unknowns.

This model assumes that each image j is shifted, rotated and scaled with respect to the master image (marked with index 'M').

Equations 2 (for standard 'master-to-slave' matching) are not analysed independently but they are included in the same system (the algorithm verifies if different pairs share the same points). Features of the master image are intended as fixed coordinates and are used to remove the rank deficiency (4 parameters) of the normal matrix of LS.

Equations (3) are used to improve the connection between the images and to strengthen block geometry. In addition, they limit the number of 'master-to-slave' features (like in a photogrammetric bundle adjustment where the number of ground control points can be reduced thanks to a set of tie points) and provide the registration parameters of images without a direct visibility to the master. The unknown pixel coordinates of these points correspond to points matched only between two or more slaves and then reprojected onto the master.

As points are extracted with the same operator (SURF) we assume that they have the same precision. This mathematical formulation fixes the points of the master image and define a reference system. The solution here proposed could be assumed as an independent model adjustment where the datum is fixed by the features extracted from the master image.

Given p image points (after data reordering, as explained in the following section), $2p$ equations can be written. The redundancy

depends on the number of images (n), i.e., four geometric parameter per every image, the coordinates of 'slave-to-slave' matches reprojected onto the master (m) and the number of 2D points (u) used as reference. The redundancy (r) of this system becomes:

$$r = 2p - 4n - 2m + 2u \qquad (4)$$

The linear system of normal equations has the following form:

$$Nx = \begin{bmatrix} A_1 & \bar{N} \\ \bar{N}^T & A_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} n_1 \\ 0 \end{pmatrix} = b \qquad (5)$$

where x_1 contains the unknown transformation parameters (a_j, b_j c_j, d_j) of image j and x_2 the image coordinates of tie points ($\underline{x}'_{ij}$, $\underline{y}'_{ij}$) projected onto the master image. N has a particular banded form, where A_1 is a hyper-diagonal matrix with sub matrices 4×4 corresponding to individual images, whereas A_1 is a $2m \times 2m$ diagonal matrix. As mentioned, the normal matrix of the linear system of normal equations a has a particular banded form and the solution $x = N^{-1}b$ is obtained with standard LS techniques.

3. AUTOMATED MATCHING OF MULTIPLE LANDSAT IMAGERY

The automated image matching procedure can be considered as a progressive process in which multiple image combinations are analysed independently, then points are reordered (i) to find corresponding points on multiple images and (ii) to run a global LS adjustment (see the previous section).

The developed methodology starts extracts corresponding points along with a feature-based matching (FBM) technique based on the SURF operator (Bay et al., 2008). The choice of this operator is motivated by its proved robustness (strongly needed in the case of automated matching where a significant percentage of outliers can be found), scale invariance (satellite images with different geometric resolutions can be potentially matched) and rotation invariance (e.g., images taken by satellites moving along different orbits and rotated footprints can be potentially matched). In addition, SURF can handle large resolution images (in this study images featuring more than 7,000×7,000 pixels have been processed) and seems more appropriate than other similar operators such as SIFT (Lowe, 2002).

The SURF operator is very used in applications requiring terrestrial images, like close-range photogrammetry and computer vision, where sub-pixel precision in image orientation was reached (Vergauwen and Van Gool, 2006; Barazzetti et al., 2010). As the descriptor is relatively new (2008), few studies with remotely sensed images have been carried out up until today. On the other hand, the technical literature reports some scientific contributions were the method was tried out with image pairs and compared to other similar operators, e.g., PCA-SIFT (Ke and Sukthankar, 2004) or GLOH (Mikolajczyk and Schmid, 2005). In these studies, sub-pixel precision and robustness against scale, position, rotation and brightness variations was achieved.

The implementation is derived from the original code available at http://www.vision.ee.ethz.ch/~surf/, and the used descriptor is a vector with 128 elements. Corresponding features can be found by simply comparing the descriptors, without any preliminary information. In addition, the *ratio test* (Barazzetti et al., 2010) with threshold T=0.75 is used to obtain more distinctive matches.

When SURF retrieves a sufficient number of image correspondences, some mismatches are often still present. To remove these outliers the procedure uses the robust estimation of a planar 2D similarity transformation between points in each image pair. Obviously, this assumes that the geometric model that connects the images includes a 2D translation, rotation and a scale factor.

The solution needs to be sought with robust techniques as they allow the detection of possible outliers in the observations. The proposed method is based on the analysis of several sets of image coordinates randomly extracted from the whole dataset. In this procedure the popular high breakdown point estimator RANSAC (Fischler and Bolles, 1981) was included.

Data (image coordinates in this case) are then clustered into regular structures ('tracks') in order to identify points visible in as many images as possible. The problem becomes similar to the extraction of tie points from aerial or close-range image blocks, where points with higher geometric multiplicity are sought to improve network geometry: a progressive check of the extracted pixel coordinates (numerical values) provides a regular structure of image coordinates, similar to the input of a standard bundle block adjustment (e.g., point label, image label, *x-y* pixel coordinates).

4. CASE STUDY

4.1 Dataset description and manual feature extraction

The dataset presented in this work is made up of a multi-temporal time series (see Table 1 for more details) including 13 Landsat-4/TM and Landsat-5/TM images (path 197, row 28) acquired over Lombardy (Northern Italy) from 1984 to 2011 and made available through the U.S. Geological Survey Earth Resources Observation and Science Center. The area is displayed in Fig. 2 in false-colour IR. All the Landsat/TM full-frames (185 km × 185 km) were processed at their original resolution (30 m for VNIR-SWIR).

Image ID	Acquisition date
1	July 25, 1984
2	January 7, 1987
3	January 23, 1987
4	February 8, 1987
5	April 29, 1987
6	July 2, 1987
7	August 3, 1987
8	September 20, 1987
9	August 13, 1988
10	August 3, 1990
11	September 16, 2003
12	August 26, 2007
13	August 21, 2011

Table 1. Landsat/TM images for the simultaneous multi-image registration example. The images of 1987 (in blue in the table) have a significant cloud cover that makes their registration really challenging.

During this year the bad weather affected the region, especially the Province of Sondrio (Northern Italy). A flood emergency started in the second half of July and lasted until the beginning of September. This makes the dataset quite challenging as the time series is made up of images collected in different seasons and years and are characterized by different cloud and snow cover.
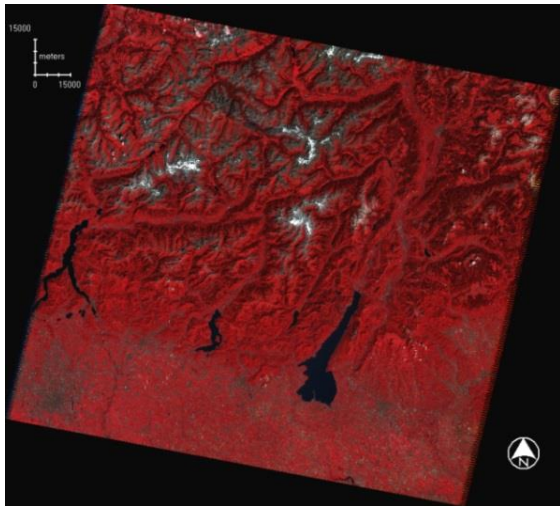
Figure 2. Landsat frame of the study area (false-colour IR).

In particular, tests carried out with ERDAS AutoSynch confirmed that standard techniques were not able to manage such a dataset, whereas the new simultaneous multi-image ('one-to-many') approach here described allowed to process all the images of the time series.

A set of multi-image features was manually measured to obtain a reference dataset. In all, 461 points were extracted in one working day, obtaining 922 equations and 78 unknowns (the same point was measured in several images). The proposed LS adjustment algorithm provided the transformation parameters of all the images and a sigma-naught of 0.57 pixels, confirming sub-pixel precision.

4.2 Automated feature matching

The automated FBM algorithm was simultaneously run for all combinations of image pairs (in this case 13 images provided 78 combinations). As previously introduced, the Landsat time series included several complex images featuring cloud and snow coverage. For this reason, different images had a different number of corresponding features with a variable geometric distribution (Fig. 3). For instance, the images acquired in summer had many corresponding features, whereas those acquired during the year of the flood (1987) had less points because of the heavy cloud coverage. In any case, all the images were included in the adjustment because the connection graph of Fig. 4 shows complete connections for all the images. Indeed, after completing the pairwise matching phase, a connection graph can be displayed to check the connection (corresponding features) between the images.
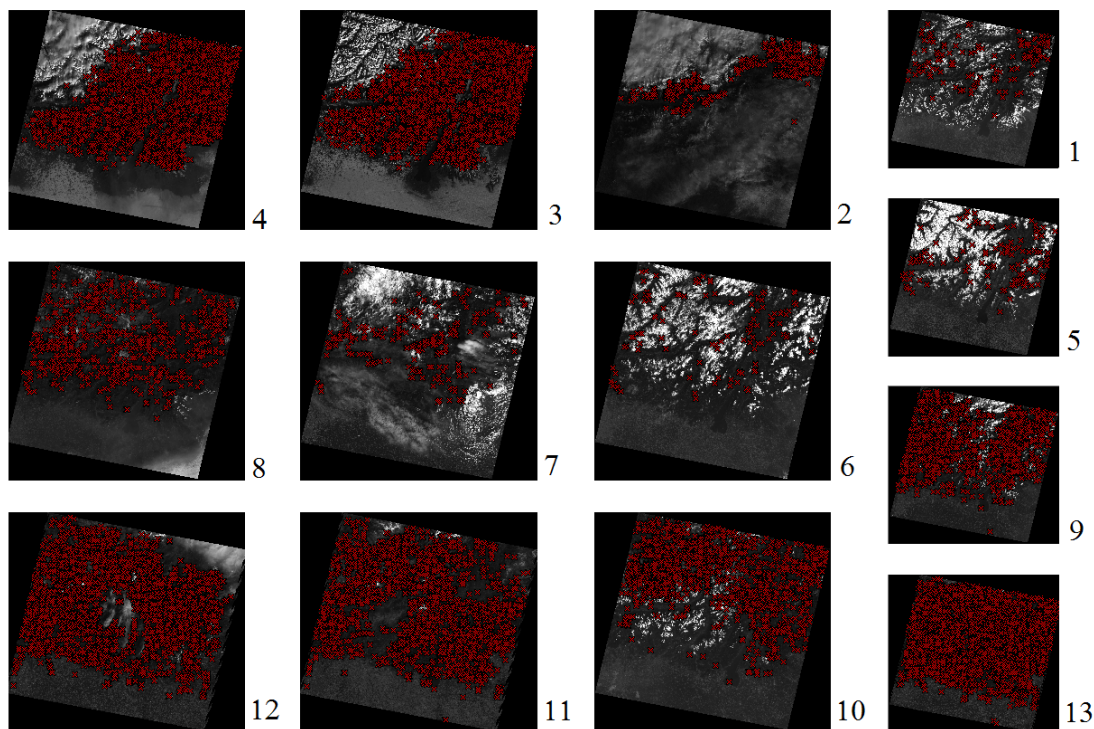

Figure 3. Automatically matched features for the whole time series.

The nodes of the graph of Fig. 4 are the satellite images and the lines their connections (meaning that homologous points were found). In our case study it is interesting to note that the graph splits the dataset into two main blocks: roughly speaking, data collected in the year of flood and the remaining ones. The first image (ID 1), acquired on July, has several connections with images 6-13 because the season is quite similar. Images 2-5 were instead collected between January and April and the connection graph clustered them together, probably on the basis of a different land-cover. In any case, all the groups are connected together and thus a global adjustment of the whole

time series is feasible. The first image of the time series was then set as reference in this experiment. Finally, LS adjustment was carried out with the implementation described in section 2. The first significant difference between manual and automated measurements concerned the number of tie points extracted: a human operator matched few tens of points whereas the automated operator provided more than 25,000 image correspondences. The balance equations *vs* unknowns gave a redundancy of 32,752 and sigma-naught of 0.73 pixel for the automatic data processing.
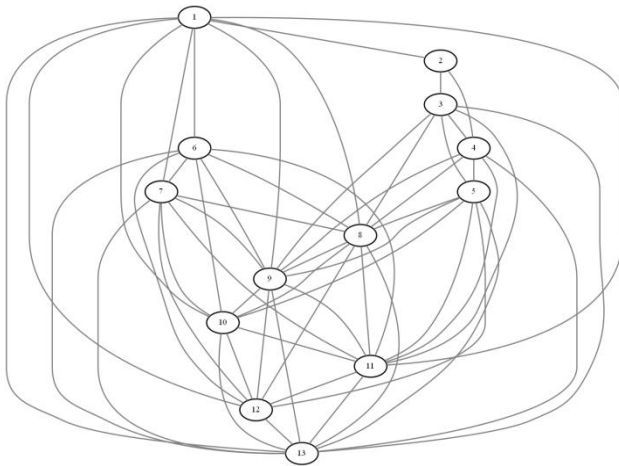
Figure 4. The connection graph after automated matching.

4.3 Accuracy evaluation: manual vs automated

As illustrated in the previous sections, average sub-pixel precision in the global registration was reached for both manual (±0.57 pixel) and automated (±0.73 pixel) measurements (sigma-naught of LS adjustment). This means that manual measurements turned out in a slightly better precision. An important result concerns the estimated scale factors and rotation angles (for both manual and automated projects): an almost unary scale factor s and a null rotation α were found for all the images, meaning that all the Landsat images used were only affected by a rigid translation. Moreover, results are comparable for both manual and automated measurements.

The results for the estimated translation vectors were instead significantly different for every image of the time series. For this reason, the differences between the translation values for manually and automatically extracted point correspondences ($\Delta_{tx} = t_{x\_man} - t_{x\_auto}$ and $\Delta_{ty} = t_{y\_man} - t_{y\_auto}$) were estimated and are shown in Fig. 5. The average and standard deviation values of the translation parameter differences resulted in -0.05±0.36 pixels (Δt_x) and -0.10±0.35 pixels (Δt_y). These sub-pixel results confirmed the consistence of automated measurements with manual data, assumed as reference in this work, and proved the correctness of the implemented matching/adjustment strategy.
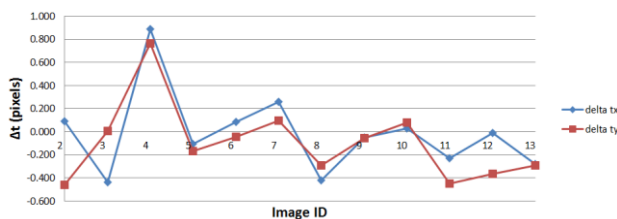

Figure 5. Comparison between results of image registration based on manual and automated measurements.

5. CONCLUSIONS AND OUTLOOK: TOWARDS MULTI-SENSOR REGISTRATION

This paper presented the first results of a new multi-image matching and adjustment methodology for the registration of satellite time series.
Image features are initially extracted in as any images as possible by checking all different image pair combinations.

Then, a registration is carried out by considering not only 'master-to-slave' features, but also 'slave-to-slave' connections into a unique rigorous Least Squares adjustment.
This method avoids standard matching between several images and a single master ('one-to-one' registration). An image without corresponding points with the 'master' can be registered in a way similar to photogrammetric bundle adjustment, where images without ground control points can be oriented by means of tie points.
The method here proposed is fully automated and very robust against outliers, as techniques for gross error detection were included.
The first experimental results proved that the method can handle complex images, notwithstanding the complexity of seasonal land-cover changes (images collected in different seasons and years), different snow and cloud cover and changes occurred in the study area.
Future developments concern the use of multi-sensor data coupled with ad-hoc procedures for feature decimation aimed at providing a more uniform point distribution (see, e.g., Gianinetto et al., 2004). A preliminary result is shown in Table 2, where the connection graph for a time series including four Landsat/TM images (30 m ground resolution) and one ASTER image (15 m ground resolution), acquired over Las Vegas.

Image ID	date	sensor	
1	November 10, 2002	ASTER (15 m)	
2	July 15, 2006	Landsat/TM	
3	June 16, 2007	Landsat/TM	
4	May 2, 2008	Landsat/TM	
5	April 2, 2009	Landsat/TM	

Table 2. ASTER and Landsat/TM images for the multi-sensor experiment.

A very important aspect for multi-sensor image matching is the selection of the spectral band to be used during the registration process: similar bands should be chosen in order to obtain a similar radiometric content and simplify the automated matching phase.
For ASTER data, spectral band nr.1 was used in this preliminary test, whereas for Landsat spectral band nr.2, as they have both collect the reflected solar radiation in the range 0.52-0.60 μm. As can be seen in Table 2, the graph reports a very strong connection between different images: there is a direct visibility between all image combinations (a maximum number of 10 combinations was reached in this case). The matching phase provided 40,576 image correspondences that resulted in 81,152 equations and 28,752 unknowns. Sigma-naught of Least Squares bundle adjustment was 0.48 pixels.
Image ID 5 was set as master and the algorithm provided a correct estimation of the scale factor for the ASTER image (0.498 was the estimated value, whereas the real value was 15 m / 30 m = 0.5).
The implemented method is usually able to extract a huge number of features visible in several images. On the other hand, all these features are not needed and can be reduced according to their multiplicity (i.e., the number of images in which the same point can be matched) and a regular grid projected onto each image. For each cell only the point with the highest multiplicity is stored. Obviously, the same point must be stored for the other images. The size of the cell depends on the geometric resolution of the images. The algorithm can

progressively decrease the size of the cell, starting from 1024 up to 512, 256, and 64 pixels, obtaining the results shown in Table 3 and the distribution shown in Fig. 6, where points are reprojected onto the master images. The estimated sigma-naught is always less than a pixel and there is also a small improvement in terms of precision according to cell size.

Cell size (pixels)	Equations	Unknowns	Sigma-naught (pixels)
64	40,196	3,518	0.53
128	21,796	1,552	0.48
256	8,716	482	0.46
512	6,098	130	0.43
1024	1,030	38	0.44

Table 3. Results with the progressive feature decimation.

These preliminary results proved how multi-sensor registration is feasible, although several aspects have to be taken into considerations. For instance, the choice of the 'best spectral band' (or band combinations) for image matching, the CPU time that still depends on the squared number of images, the use of fast strategies for descriptor comparison, the implementations of other geometric models, and the selection of the master image according to the topologic structure of the connection graph. Most of the listed points will be considered in future developments to implement a multi-image multi-sensor registration algorithm.
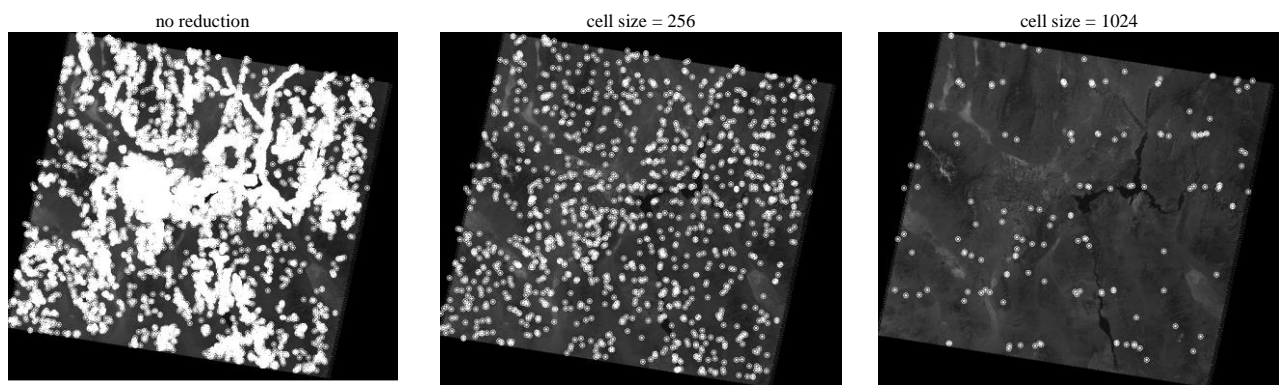
ACKNOWLEDGEMENTS

Figure 6. Features reprojected onto the master after Least Squares adjustment incorporating feature decimation.

REFERENCES

Bay H., Ess A., Tuytelaars T., Van Gool L. (2008). SURF: Speeded Up Robust Features. Computer Vision and Image Understanding, 110(3): 346-359.

Barazzetti L., Remondino F., Scaioni, M. (2010). Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. The Photogrammetric Record, 25(132): 356-381.

Castro E.D., Morandi C. (1987). Registration of translated and rotated images using finite Fourier transform. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9:700-703.

Fischler M. A., Bolles R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, 24: 381-395.

Gianinetto M., Scaioni M., Borgogno Mondino E., Tonolo G.F. (2004). Satellite images geometric correction based on non-parametric algorithms and self-extracted GCPs. In: Proc. IGARSS '04, Anchorage, Alaska-USA, 20-24 Sept., Vol. 6, pp. 3755-3758.

Gianinetto M., Scaioni M. (2008). Automated Geometric Correction of High-Resolution Pushbroom Satellite Data. Photogrammetric Engineering & Remote Sensing, 74(1): 107-116.

Gianinetto M. (2012). Automatic Co-Registration of Satellite Time Series. The Photogrammetric Record, 27(140): 462-470.

Gosthtasby A. (2005). 2-D and 3-D image registration. John Wiley & Sons, Hoboken, NJ, U.S.A., 258 pages.

Hartley R., Zisserman A. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518.

Ke Y., Sukthankar R. (2004). PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. Computer Vision and Pattern Recognition.

Kraus K. (2007). Photogrammetry: Geometry from Images and Laser Scans. Walter de Gruyter, 459 pages.

Le Moigne J., Netanyahu N.S., Eastman R.D. (2011). Image Registration for Remote Sensing. Cambridge University Press, U.K., 484 pages.

Lowe D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2): 91-110.

Mikolajczyk K., Schmid C. (2005). A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(27): 1615-1630.

Pluim J.P.W., Maintz J.B.A., Viergever M.A. (2001). Mutual information matching in multiresolution contexts. Image and Vision Computing, 19: 45–52.

Poli D., Toutin T. (2012). Review of developments in geometric modelling for high resolution satellite pushbroom sensors. The Photogrammetric Record, 27(137): 58–73.

Pratt W.K. (1991). Digital Image Processing. 2nd ed., John Wiley & Sons, Inc. New York, NY, USA.

Price K.E. (1985). Relaxation matching techniques - a comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7:617–623.

Townshend J.R., Justice C.O., Gurney C., McManus J. (1992). The impact of misregistration on change detection. IEEE Transactions on Geosciences and Remote Sensing, 30(5): 1054-1060.

Tucker C.J., Grant D.M., Dykst J.D. (2004). NASA's Global Orthorectified Landsat Data Set. Photogrammetric Engineering and Remote Sensing, Vol. 70, No. 3, pp. 313–322

Vergauwen M., Van Gool L. (2006). Web-Based 3D Reconstruction Service. Machine Vision Applications, 17: 411-426.