AUTOMATED DSM BASED GEOREFERENCING OF CARTOSAT-1 STEREO SCENES

Pablo d'Angelo[a], Peter Schwind[a], Thomas Krauss[a], Frithjof Barner[b] and Peter Reinartz[a]

[a]German Aerospace Center (DLR), Remote Sensing Technology Institute, D-82234 Wessling, Germany
email: {Pablo.Angelo, Peter.Schwind, Thomas.Krauss, Peter.Reinartz}@dlr.de

[b]Euromap GmbH, Kalkhorstweg 53, 17235 Neustrelitz, Germany
email: barner@euromap.de

KEY WORDS: Georeferencing, Digital elevation models (DEM), Digital photogrammetry, Feature extraction, Stereo, CARTOSAT-1

ABSTRACT:

High resolution stereo satellite imagery is well suited for the creation of digital surface models (DSM). A system for highly automated and operational DSM and orthoimage generation based on CARTOSAT-1 imagery is presented, with emphasis on fully automated georeferencing. The proposed system processes level-1 stereo scenes using the rational polynomial coefficients (RPC) universal sensor model. The RPC are derived from orbit and attitude information and have a much lower accuracy than the ground resolution of approximately 2.5 m. In order to use the images for orthorectification or DSM generation, an affine RPC correction is required. This requires expensive and cumbersome GCP acquisition. In this paper, GCP are automatically derived from lower resolution reference datasets (Landsat ETM+ Geocover and SRTM DSM). The traditional method of collecting the lateral position from a reference image and interpolating the corresponding height from the DEM ignores the higher lateral accuracy of the SRTM dataset. Our method avoids this drawback by using a RPC correction based on DSM alignment, resulting in improved geolocation of both DSM and ortho images. The proposed method is part of an operational CARTOSAT-1 processor at Euromap GmbH for the generation of a high resolution European DSM. Checks against independent ground truth indicate a lateral error of 5-6 meters and a height accuracy of 1-3 meters.

1 INTRODUCTION

In May 2005 India launched its IRS-P5 satellite with the CARTOSAT-1 instrument which is a dual-optics 2-line along-track stereoscopic pushbroom scanner with a stereo angle of $31°$ and the very interesting resolution of 2.5 m. The CARTOSAT-1 high resolution stereo satellite imagery is well suited for the creation of digital surface models (DSM). In this paper, a system for highly automated DSM generation based on CARTOSAT-1 stereo scenes is presented. More details about CARTOSAT-1 are given in (Srivastava et al., 2007).

CARTOSAT-1 stereo scenes are provided with rational polynomial functions (RPC) sensor model (Grodecki et al., 2004), derived from orbit and attitude information. The RPC have a much lower accuracy than the ground resolution of approximately 2.5 m. Subpixel accurate ground control points (GCP) have been used in previous studies to estimate bias or affine RPC correction parameters required for high quality geolocation of HRSI images (Lehner et al., 2007). Such highly accurate GCP are usually derived from a DGPS ground survey or high resolution orthoimages and digital elevation models. For large scale and continent wide processing, establishing a highly accurate GCP database with the density required for processing the relatively small CARTOSAT-1 scenes (900 km^2) requires significant resources. For near real-time applications such as disaster assessment tasks in remote regions, highly accurate GCP data is often not available.

We propose the use of widely available lower resolution satellite data, such as the Landsat ETM+ and SRTM DSM datasets as a reference for RPC correction. The accuracy of these datasets is low compared to the high resolution CARTOSAT-1 images. The absolute lateral error of ETM+ Geocover is 50 m (LE90), the absolute lateral error of SRTM is between 7.2 m and 12.6 m (LE90, depending on the continent), with an absolute height error of 4.7 m to 9.8 m (Rodriguez et al., 2005). The traditional method of collecting lateral position from a reference image and interpolating the corresponding height from the DEM ignores the higher lateral accuracy of the SRTM dataset. Our method avoids this drawback by using a RPC correction based on DSM alignment, resulting in improved geolocation of both generated DSM and ortho images.

Digital surface models are derived from dense stereo matching and forward intersection and subsequent interpolation into a regular grid. The first part of the paper describes the process used for DSM generation, with specific emphasis on the georeferencing. The second part evaluates the processor using four CARTOSAT-1 stereo pairs.

2 CARTOSAT-1 STEREO PROCESSOR

2.1 Stereo matching

Hierarchical intensity based matching is used for matching the stereo pairs and the reference image. It consists of two major steps, hierarchical matching to derive highly accurate tie points and dense, epipolar based stereo matching.

The initial matching step uses a resolution pyramid (Lehner et al., 1992; Kornus et al., 2000) to cope with large stereo image distortions originating in carrier movement and terrain. Large local parallaxes can be handled without knowledge of exterior orientation. The selection of pattern windows is based on the Foerstner interest operator which is applied to one of the stereo partners. For selection of search areas in the other stereo image(s), local affine transformations are estimated based on already available tie points in the neighbourhood (normally from a coarser level of the image pyramid). Tie points with an accuracy of one pixel are located via the maximum of the normalized correlation coefficients computed by sliding the pattern area all over the search area. These approximate tie point coordinates are refined to subpixel accuracy by local least squares matching (LSM). The number of points found and their final (subpixel) accuracy achieved depend mainly on image similarity and decrease with increasing

stereo angles or time gaps between imaging. Strict thresholds on correlation coefficient and bidirectional matching differences are used to select reliable and highly accurate stereo tie points.

An epipolar stereo pair with epipoles corresponding to the image columns is generated by aligning the columns of the Fore image with the Aft image, using highly accurate matches from the pyramidal matching step. Dense stereomatching is performed on the epipolar images, using either a regiongrowing LSM matching (Heipke et al., 1996) or semiglobal matching (SGM) (Hirschmüller, 2008). The regiongrowing matching has a higher accuracy and less outliers in flat, textured regions, but cannot properly handle the discontinuities at man made objects or forest boundaries that are clearly visible in the 2.5 m CARTOSAT-1 imagery. The semiglobal matching approach avoids these drawbacks, but results in more outliers in image areas with low contrast and different appearance in both images, such as large agricultural areas. Extending SGM to handle these areas properly is ongoing work.

2.2 Georeferencing using stereo scenes

After high quality tie points between the stereo pair have been found, they are transferred to the lower resolution reference image using local least squares matching. Due to strong radiometric differences between the ETM+ and the CARTOSAT-1 imagery, only a small fraction of the stereo tie points can be matched reliably between all three images. Three-dimensional GCP are found by bilinear interpolation of the SRTM DSM at the geographic positions found in the ETM+ reference scene. These GCPs are called Image GCP in the following sections. The Image GCPs are used to estimate affine RPC correction parameters for each of the stereo images.

After the alignment based on ETM+ location and SRTM height reference data, forward intersection residuals are significantly improved, but the lateral and vertical accuracy is still limited by the ETM+ reference.

The higher lateral accuracy of the SRTM dataset is used in a second RPC correction step. A 3D point cloud is calculated by forward intersection of a subset of the stereo tie points. The point cloud is aligned to the SRTM DSM using 3D least squares matching. The estimated 3D affine transformation is then used to estimate affine RPC correction parameters based on the SRTM geolocation. This procedure is described in detail in Section 2.2.1.

As CARTOSAT-1 requires an affine RPC correction, well distributed GCPs are essential for high quality results. The 3D alignment process works well in areas where significant terrain is available throughout the whole scene. Scenes that contain no usable relief for DSM alignment can only be corrected by using the Image GCP. If a scene contains enough well distributed relief, the DEM GCPs are used (see Section 2.2.1). Problems also occur for hilly scenes that contain large flat areas, as the 3D matching does not provide lateral constraints in flat areas. This is similar to using an uneven distribution of ground control points. For these scenes, the lateral shift between ETM+ and SRTM is estimated in areas with high relief and used to correct the ETM+ based GCP in flat areas, establishing consistent GCPs in both flat and high relief areas. The affine RPC correction is then estimated using both the corrected ETM+ based GCPs and the SRTM based GCPs, cf. Section 2.2.2.

2.2.1 RPC correction by DSM alignment: A 3D point cloud is calculated by forward intersection of a subset of the stereo tie points. The point cloud is aligned to the SRTM DSM. It is assumed that the height z_i of a point P_i located at (x_i, y_i, z_i) equals

the reference DSM height $h_D(x_i, y_i)$ at the corresponding position (x_i, y_i):

$$h_D(x_i, y_i) = z_i \qquad (1)$$

A 3D affine transformation is used to align the initial stereo point cloud to the SRTM DSM:

$$p_{ti} = A p_i \qquad (2)$$

where $p_i = (x_i, y_i, z_i, 1)$ is the original point, A is a 3×4 matrix, and $p_{ti} = (x_{ti}, y_{ti}, z_{ti})$ is the transformed point.

The affine transformation matrix A is estimated using an iterative least mean squares algorithm. Using Eq. 1 and 2, the following observation equation is obtained:

$$v_i = h_D(x_{ti}, y_{ti}) - z_i. \qquad (3)$$

Since the model is non-linear, the solution is obtained iteratively. An identity transform is used as initial approximation, since the stereo points are not far from the reference. It is likely that the stereo point cloud, and to a smaller extend the DSM contains outliers, which cannot be handled by a standard least mean squares algorithm. After the initial estimation, points with a residual larger than 3 times the standard deviation are removed and a new transformation is estimated. This procedure is repeated until less than 0.3 % outliers are detected and the squared sum of the outlier residuals accounts for less than 5 % of the squared sum of all residuals.

The estimated affine transformation could be used to align the final DSM to the SRTM reference and thus improve its accuracy. Orthoimages would however still be limited by the ETM+ accuracy. It is desirable to include the correction in the RPC models, too. This is done by using the aligned stereo points as GCPs for a second RPC correction which yields the affine RPC correction parameters. These GCPs are referred to as DEM GCPs in this paper.

2.2.2 Using DSM and Reference Image: Matching the 3D point cloud with the SRTM DEM using the 3D least squares matching can only succeed if there are distinct features in the terrain being matched. On flat terrain this matching method will not lead to an improved lateral accuracy.

To overcome this problem, the GCPs in areas with high relief, where the 3D alignment process works accurately, are used to estimate the shift between ETM+ and SRTM. Areas suitable to extract the GCPs used here are identified by applying an interest area operator similar to the one used by the SIFT algorithm (Lowe, 1999) to the SRTM DSM.

The DSM is convolved repeatedly with a Gaussian filter, creating the scale-space $L(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (4)$$

These Gauss-filtered images where the standard deviation differs by a factor k are subtracted from each other to create the Difference of Gaussians (DoG) images $D(x, y, \sigma)$:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad (5)$$

Extrema are detected in this DoG pyramid by comparing every pixel to the eight neighbouring pixels and the nine pixels in the scales above and below. If a pixel is larger or smaller than all of its neighbours, it is accepted as a preliminary keypoint candidate.

Next, all keypoint locations are interpolated with subpixel accuracy, using an iterative method (Brown and Lowe, 2002). As the

position is interpolated in all three dimensions of the scale space, this also affects the size of a detected area. After the interpolation, two additional checks are performed to remove unstable keypoints. First the value at the extremum $D(\hat{x})$ is computed. Keypoints with a value below a certain threshold are eliminated, thereby removing points with low contrast. Then, points lying on ambiguous edges rather than corners are removed, making use of the Hessian matrix H computed at the keypoint location:

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \tag{6}$$

The derivatives D_{xx}, D_{yy} and D_{xy} are determined by calculating the differences between neighbouring points. Rather than solving the Eigenvalue problem, keypoints are selected from the trace $Tr(H)$ and the determinant $Det(H)$ by requiring

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{7}$$

where r is usually set to 10 (Lowe, 2004).

The region around each of the keypoints which would be used by the SIFT operator to generate a descriptor for matching, is instead used to create a circular mask. This way regions deemed suitable for the 3D point cloud alignment (mountains, high hills) can be distinguished from regions unsuitable for this purpose (lowlands, shallow hills) (see Figure 1). Thus, the image mask can be used to filter out stereo tie points lying in areas with only minor height differences. It is also used to decide which of the three RPC correction methods should be used.
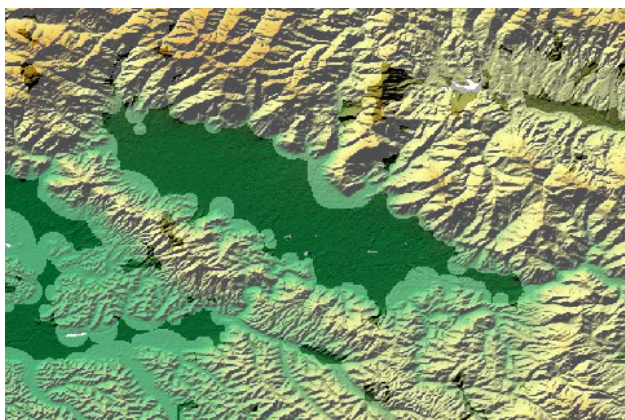

Figure 1: Detected regions suitable for DSM alignment (Background: shaded SRTM). Bright areas indicate high relief areas

The remaining stereo tie points are processed using a forward intersection and are aligned to the DEM, cf. Section 2.2.1. Since alignment of the 3D point cloud was shown to work accurately in areas with significant terrain (d'Angelo et al., 2008), the translation of the stereo tie points can be assumed to describe the displacement between the ETM+ and the SRTM. Due to this the average shift between the resulting points and the initial image stereo tie points can be used to correct the lateral shift of the original stereo tie points.

As the transformation from the Landsat ETM+ to the SRTM DSM can sometimes be more complicated than a simple displacement, the standard deviation of the shift is computed. In case of a high standard deviation no uniform translation was found and this method is not expected to significantly improve the performance of the RPC correction.

Finally, these corrected tie points together with the previously generated SRTM GCPs are used to estimate the new RPC correction parameters. Because of their generally higher accuracy, the SRTM GCPs are usually preferred for the RPC correction. That is why they are used when available, while the GCPs obtained by matching with the reference image are only used in areas where no SRTM GCPs are available. To ensure an even distribution of GCPs over the scene they are thinned out using a grid. This is necessary as an uneven distribution could result in a region of the scene with a disproportionately high number of GCPs outweighing the rest of the GCPs, leading to distorted RPC correction parameters.

2.3 DSM and Orthoimage

After forward intersection of the dense stereo matching result and subsequent reprojection into UTM, the forward intersected points still contain a small amount of blunders due to matching errors in regions with sparse texture. To eliminate gross outliers, a reference check against the SRTM DSM is performed. All points whose height deviates more than 75 m from the SRTM are removed. A second outlier check is done by dividing the data set in small, overlapped tiles. The minimum and maximum points in each tile are tested using a critical threshold which is calculated from the height points inside the tile by means of Grubbs' test algorithm (Grubbs, 1969). In each iteration one point, i.e., either maximum or minimum points, is tested and if the point is rejected by the test, the next extrema is validated until stability is reached, i.e., neither minimum nor maximum are rejected.

The resulting DSM contains holes in areas where the matching failed or outliers were removed. These holes are filled with SRTM data using the delta surface fill method (Grohman et al., 2006). Orthorectified images are produced using the affine RPC correction and the generated DSM.

3 EVALUATION

To demonstrate the improved performance of the RPC correction based on DEM and combined GCPs compared to using only image GCPs, the three methods are analysed. For this purpose four CARTOSAT-1 scenes of Catalonia are processed (see Figure 2). Scene 114208 was part of the Cartosat scientific assessment programme, while the remaining 3 scenes have been provided by Euromap. In addition to that several orthoimages with a resolution of 0.5 m and two DTM with 15 m resolution have been provided by the Institut Cartographic de Catalunya (ICC) and are only used as ground truth during the evaluation.

The scenes are mostly cloudless. The scenes located near the coastline contain both flat areas along the coast as well as the Montseny mountain range with peaks of over 1600 meters. As shown in Table 1, the scenes were acquired early in the year, leading to large shadows in the mountainous areas.

Evaluation of the accuracy relative to the ground truth scenes is realized in four main steps:

1. Independent checkpoints are obtained by manually matching the Aft images of the four scenes with the ground truth orthoimages. The height of the points in the ground truth images is estimated using two reference DTMs.

2. The corresponding points in the Fore images are matched automatically, based on the already obtained matches of the stereo matching phase of the CARTOSAT-1 stereo processor.
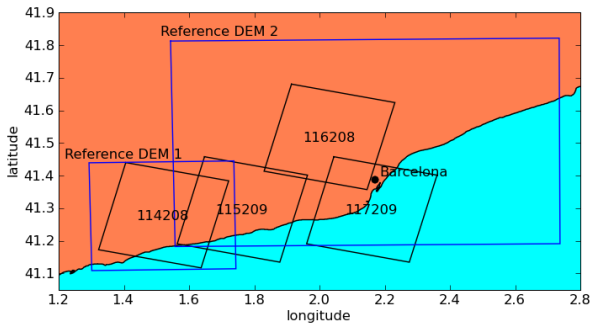
Figure 2: Geographic layout of evaluated CARTOSAT-1 scenes and Reference DEMs used for verification.

Scene	Imaging date
114208	01 Feb. 2006
115209	16 Feb. 2008
116208	05 March 2008
117209	25 Jan. 2008

Table 1: CARTOSAT-1 scenes evaluated in this paper

3. These stereo points are processed during the forward intersection step using the three different RPC correction methods based on Image GCPs, DEM GCPs and the combination of both (see Section 2.2).

4. The displacements of the computed intersection points relative to the independent checkpoints measured in the orthoimages in Step 1 is determined.

3.1 Matching

As can be seen in Table 2 a sufficient number of points was matched in all four scenes. The differences in the amount of matches can be explained easily by the proportion of water in the imaged areas. Apart from the number of matches also the height differences to the reference DTM after the forward intersection are shown. The relatively high standard deviations here are caused by the suboptimal conditions for image matching (low sun angle, mountainous terrain with vegetation). Scene 117209 contains the city of Barcelona with many high buildings, resulting in a larger difference between the bare ground reference DTM and the generated surface model.

3.2 Georeferencing

Due to space limitations, only results of scene 114208 are discussed here, as similar results were obtained for the other three tested scenes.

The displacements of the GCPs computed by the three tested methods in scene 114208 relative to the ground-truth images can be seen in Figure 3. As expected, the method based solely on image GCPs matched in ETM+ images achieves the lowest accuracy

Scene	Number accepted points (Mio.)	Height difference to DTM Mean(m)	σ(m)
114208	24.1	-0.13	3.62
115209	20.0	-1.90	6.55
116208	31.3	-1.52	5.22
117209	9.2	-4.63	6.80

Table 2: Accepted object points statistics.

out of the three tested methods. The two other methods perform pretty similar which can be attributed to the fact that the evaluated area offers many terrain features suitable for the computation of DEM GCPs with high accuracy. Since the third method based on image and DEM GCPs favours DEM GCPs when available, its performance does not deviate much from the DEM GCP approach.
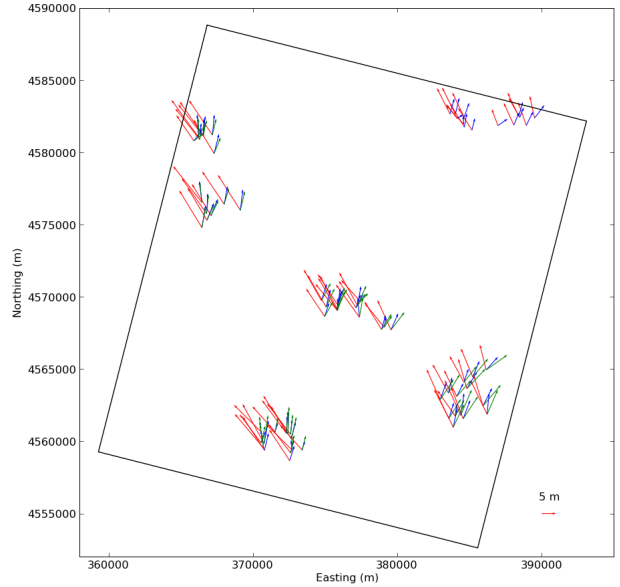

Figure 3: Lateral shifts of the three tested correction methods in scene 114208, computed using 68 manually measured independent checkpoints. (red: Image GCP; blue: DEM GCP; green: Combined GCP; arrow lengths are scaled by a factor 200)

The observations made in this scene are confirmed in the more detailed statistics shown in Table 3. The image based GCPs have the highest lateral displacements out of all tested methods (11.94m) while the other two approaches have similar displacements (DEM GCPs: 5.70 m, Combined GCPs: 6.91 m). The displacement in height varies only little and is low in all three cases (-0.16 m to 1.59 m). Considering that the ETM+ has a resolution of 15 m, the 11.94 m lateral displacements can be considered a good result, indicating subpixel accuracy. The even lower displacements for the other two approaches however show that the image based approach is outperformed easily by the two methods making use of the SRTM GCPs.

3.2.1 RPC correction using DSM and reference image: As the performance of the image and DEM GCP based approaches has already been discussed in detail before (d'Angelo et al., 2008), the performance of the combined approach in special will be discussed hereafter.

The displacements of the image GCPs lying within the generated image mask relative to the positions computed by the 3D cloud alignment are shown in Figure 4. The average shift calculated for these points was -9.14 m in x and 4.98 m in y, with a standard deviation of 3.63 m in x and 3.87 m in y. The shifts illustrated in Figure 4 as well as the computed standard deviation indicate that the transformation between the ETM+ and the SRTM in this case can be described by a simple translation, when considering the resolution of the ETM+ reference image. Tests with other scenes have sometimes shown more complex displacement patterns, probably caused by the coarser DEM used for ETM+ orthorectification. Unfortunately for none of these cases ground

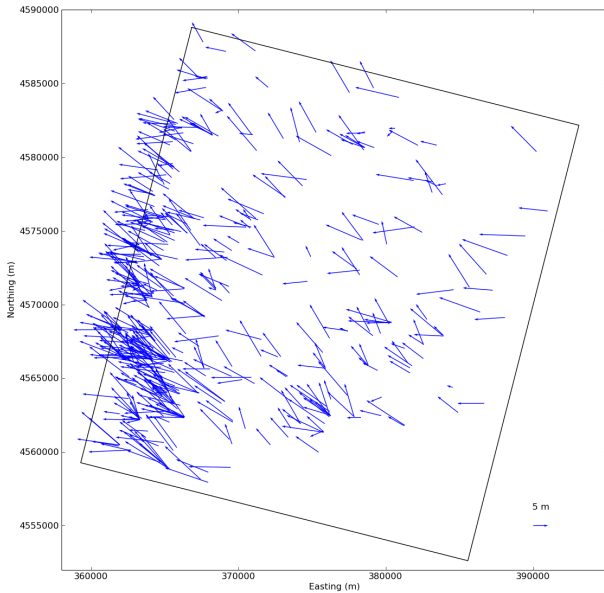truth was available for a comprehensive evaluation at the time of writing.

Figure 4: Shifts between ETM+ reference image and SRTM reference DEM (arrow lengths are scaled by a factor 200).

Even though the processed scene was not ideal for the application of the combined approach, the DEM computed using the suggested algorithm is still closely aligned to the reference ICC DTM (see Figure 5). The generated DEM at most deviates by about 15 m from the reference DTM.
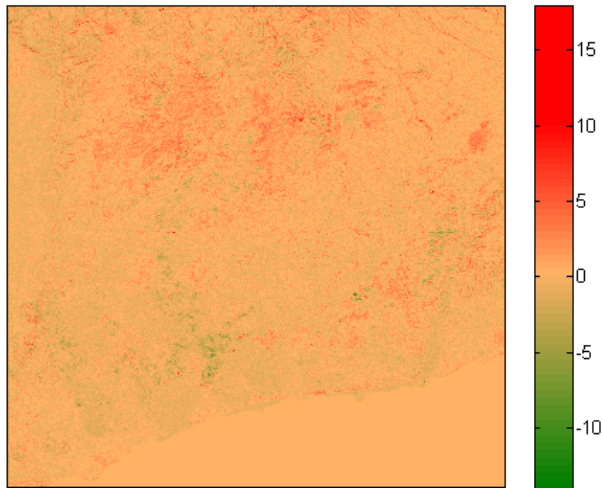

Figure 5: Height difference in meters between the reference ICC DTM and the CARTOSAT-1 DEM generated using the combined approach of scene 114208

3.3 Mosaicking

The four evaluated scenes together with six additional scenes were used to create orthoimage (Figure 6(a)) and DEM (Figure 6(b)) mosaics of the area around Barcelona. A more detailed view of the overlap between scene 116208 and 117209 is shown in Figure 6(c). Using SIFT matching an average lateral shift of -0.04 m in x and 0.31 m in y with a standard deviation of 0.78 m

	Image GCP	DEM GCP	Combined
Lateral shift (m):			
Mean	11.94	5.70	6.91
σ	2.27	1.33	1.52
Min	6.10	2.56	4.00
Max	16.08	8.16	9.91
Height shift (m):			
Mean	1.18	-0.16	1.59
σ	1.86	1.82	2.17
Min	-8.73	-10.05	-6.19
Max	4.62	3.28	6.48

Table 3: Accuracy of the different approaches, evaluated with 68 independent checkpoints

in x and 0.83 m in y was measured for these two, independently processed scenes, meaning the scenes are aligned with subpixel accuracy. As can be seen in Figure 6(d) the created Cartosat-1 DEM shows much more details than the reference SRTM C Band DSM. For example buildings, which are recognizable to some extent in the Cartosat-1 DEM, are missing completely in the SRTM DSM.

4 CONCLUSIONS

A CARTOSAT-1 processor for highly automated and operational DSM and orthoimage generation was presented. Three different RPC correction approaches implemented in the processor have been evaluated in this paper. It is shown that a RPC correction approach based on matching points with a Landsat ETM+ can be improved substantially by aligning a 3D point cloud generated through forward intersection to a SRTM DSM. To make sure that such an approach also works in scenes where only few prominent terrain features are available, the shift between reference image and reference DEM (SRTM) is estimated and used to correct the image based GCPs. Using these improved methods it was shown, that relative to independent ground truth images, an accuracy of 5-6 meters on ground and 1-3 meters in height is achieved.

The presented method is part of an automated DSM production chain used by Euromap GmbH for the generation of a high resolution European DSM.

In a future evaluation it would be interesting to analyse a scene more suitable for the new RPC correction mode based on correcting the shift between ETM+ and SRTM data. Unfortunately, no reference scene with ground truth for such a case was available at this time.

REFERENCES

Brown, M. and Lowe, D., 2002. Invariant features from interest point groups. In: British Machine Vision Conference, pp. 656–665.
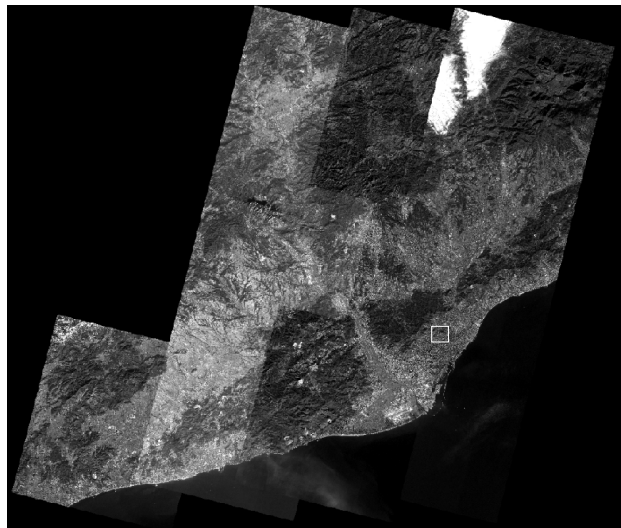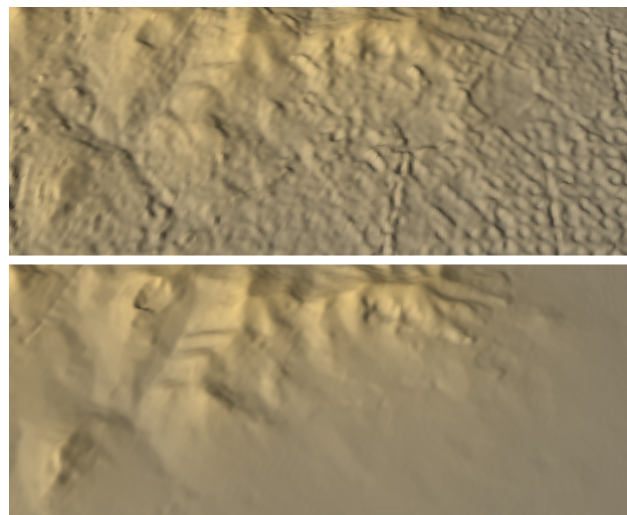
(a) Orthoimage mosaic without radiometric corrections

(b) DSM mosaic

(c) Overlap between scenes 116208 and 117209

(d) The generated Cartosat-1 DSM (top) compared to the corresponding SRTM C Band DSM

Figure 6: Mosaic of the four evaluated scenes and additional surrounding scenes

d'Angelo, P., Lehner, M., Krauss, T., Hoja, D. and Reinartz, P., 2008. Towards automated DEM generation from high resolution stereo satellite images. In: ISPRS Conference 2008, Vol. XXXVII, pp. 1137 – 1342.

Grodecki, J., Dial, G. and Lutes, J., 2004. Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conf. Proc.

Grohman, G., Kroenung, G. and Strebeck, J., 2006. Filling SRTM voids: The delta surface fill model. Photogrammetric Engineering and Remote Sensing 72(3), pp. 213–216.

Grubbs, F., 1969. Procedures for detecting outlying observations in samples. Technometrics 11(1), pp. 1–21.

Heipke, C., Kornus, W. and Pfannenstein, A., 1996. The evaluation of MEOSS airborne 3line scanner imagery processing chain and results. Photogrammetric Engineering and Remote Sensing 62(3), pp. 293–299.

Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), pp. 328 – 341.

Lehner, M., Müller, R., Reinartz, P. and Schroeder, M., 2007. Stereo evaluation of CARTOSAT-1 data for french and catalonian test sites. In: ISPRS Hannover Workshop 2007: High Resolution Earth Imaging for Geospatial Information, Vol. XXXVI.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: International Conference on Computer Vision ICCV, Corfu, pp. 1150–1157.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 20, pp. 91–110.

Rodriguez, E., Morris, C., Belz, J., Chapin, E., Martin, J., Daffer, W. and Hensley, S., 2005. An assessment of the SRTM topographic products. Technical Report Technical Report JPL D-31639, Jet Propulsion Laboratory, Pasadena, California.

Srivastava, P., Srinivasan, T., Gupta, A., Singh, S., Nain, J., Amitabh, P. S., Kartikeyan, B. and Gopala, K. B., 2007. Recent advances in CARTOSAT-1 data processing. Proceedings ISPRS Hannover Workshop 2007: High Resolution Earth Imaging for Geospatial Information.