

A SEMIAUTOMATIC APPROACH FOR GENERATION OF SITE MODELS FROM CARTOSAT-2 MULTIVIEW IMAGES

Archana Mahapatra, Sumit Pandey, D Sudheer Reddy, P S Subramanyam, B. K. Das, P V Radhadevi, J Saibaba, Geeta Varadan

Advance Data Processing Research Institute, Department of Space,
203 Akbar Road, Secunderabad-500009, INDIA,
archaana@adrin.res.in

Commission III, WG III/4, III/5

KEY WORDS: Site Model, Physical Sensor Model, Relative orientation, conjugate points, edge extraction, normalized DSM

ABSTRACT:

In the last decade there has been a paradigm shift in creating, viewing and utilizing geospatial data for planning, navigation and traffic management of urban areas. Realistic, three-dimensional information is preferred over conventional two dimensional maps. The paper describes objectives, methodology and results of an operational system being developed for generation of site model from Cartosat-2 multiview images. The system is designed to work in operational mode with varying level of manual interactivity. A rigorous physical sensor model based on collinearity condition models the 'step n stare' mode of image acquisition of the satellite. The relative orientation of the overlapping images is achieved using coplanarity condition and conjugate points. A procedure is developed to perform digitization in mono and stereo modes. A technique for refining manually digitized boundaries is developed. The conjugate points are generated by establishing a correspondence between the points obtained on refined edges to analogous points on the images obtained with view angles ± 26 deg. It is achieved through geometrically constrained image matching method. The results are shown for a portion of multi-view images of Washington City obtained from Cartosat-2. The scheme is generic to accept very high resolution stereo images from other satellites as input.

1. INTRODUCTION

A site model represents both the natural and man-made features such as terrain, road, and buildings to sufficient level of detail so that it can be considered as a true model of the area under consideration. Buildings are important objects of any 3D city model. The algorithms for fully automatic extraction of buildings are not matured enough as a result many researchers in this field opt for semi automatic methods (Gruen 99).

Building extraction requires representation of roof structures. A finer model represents roof details, overhangs and includes realistic texture of the building (Ulm and Poli, 2006). It can also have architectural details of the building. In a coarse model, detection of the buildings is only possible. A DSM rendered with texture may be considered as coarse city model (Kraub, 2008).

The inputs required for generating finer details includes aerial images, terrestrial images, LIDAR measurements and building plans (Vosselman et al, 2001). The geometric resolution and radiometric quality of the images are important as it should be possible to identify and delineate features.

Aerial images have the advantage of very high geometric resolution and low noise levels. In the last few years, there has been considerable improvement in the technology to provide better resolution images from space platforms. The images from IKONOS, Quickbird, Worldview-1/2, Geoeye-1, and Cartosat-2 have geometric resolution less than one meter. These geometric resolution levels are not sufficient to generate finer models, however extraction and representation of major features are still possible.

The objective of this work is to develop a semiautomatic system to generate site models from satellite stereo images/multiview images in operational mode. At present the focus is mainly on

the extraction and representation of building flat roof tops. It is not envisaged to generate a photorealistic site model, reconstruction of complex building roof structures and individual representation of trees.

The paper describes in brief the imaging geometry of Cartosat-2 and physical sensor model developed, followed by details of computation of relative orientation parameters and comparative analysis of rational function model with physical sensor model. A method of refining digitized edges with the help of canny edges is briefed. The semiautomatic mode of data capture and image matching techniques are described subsequently. A section is devoted to explain generation of DSM and DTM. Finally results are presented for a portion of the Washington city.

2. METHODOLOGY

2.1 Cartosat-2 Imaging Geometry

India's highest resolution imaging satellite, Cartosat-2, was launched in 2007. The spatial resolution of better than one meter in panchromatic band is achieved through 'step n stare' mechanism of image acquisition. In this image acquisition process, the effective ground velocity is reduced by continuously steering camera in the direction opposite to spacecraft motion. Reduction of ground trace velocity in turn improves the spatial resolution of the image in along-track direction. To reduce the pixel cross talk and improve the image quality, the CCD arrays have staggered configuration in the focal plane. Even and odd pixels of a line are recorded by the two detector arrays that are separated by 35μ in the focal plane. It means that the adjacent pixels of a line are not imaged at same instant. If all odd numbered pixels of a line are imaged at instant t , the even numbered pixels will be imaged at instant represented by $t + 5 \times$ integration time.

The spacecraft has three modes of image acquisition; these modes are termed as spot, paintbrush and multiview modes. The spot mode covers a swath of 9.6 km with strip length varying from 6 to 290 km. The paintbrush mode provides a wider combined swath by imaging adjacent strips from same orbit. In multiview mode, the same area is imaged from two or three different view angles from the same orbit as shown in Fig. 1. This mode is useful for computation of height of the objects. However, due to continuous variation of pitch rate, the along-track resolution and the base to height ratio of multiview image acquisition varies for each imaged line. Due to these factors modelling the imaging geometry of spacecraft becomes complex. A physical sensor model is developed that takes into account the dynamic nature of imaging process of Cartosat-2.

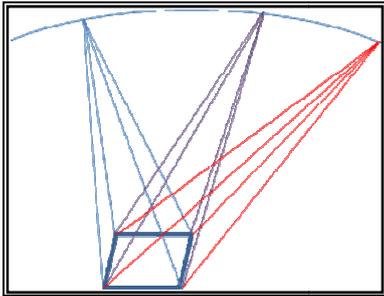


Fig. 1: Cartosat-2 Multiview Imaging Mode

2.2 Physical Sensor Model for Cartosat-2

The Cartosat-2 spacecraft is equipped with satellite positioning system, star sensors and gyros to provide the position and orientation information at regular time intervals. The physical sensor model utilizes this information in a systematic and coherent manner. The model does not approximate the shape of the orbit. The osculating nature of the orbit is accounted by converting the position and velocity parameters to slowly varying Keplerian elements, which are interpolated to know the position at the time of imaging. The orientation information is available as set of quaternions, which are converted to Euler angles. The residual orientation error is modelled as bias in roll, pitch, and yaw over a short segment of imaging. In case, precise control points are not available the positional accuracy is improved using control points identified from Enhanced Thematic Mapper (ETM) orthoimages and SRTM DEM.

The model is based on well known collinearity condition which states that the object position, image position and the perspective centre lie on a straight line at the time of imaging. Equation (1) represents the collinearity condition in mathematical form.

$$\begin{bmatrix} x \\ y \\ -f \end{bmatrix} = s.M. \begin{bmatrix} X - X_p \\ Y - Y_p \\ Z - Z_p \end{bmatrix} \quad (1)$$

In equation (1), x and y represent the image plane coordinates, f is the effective focal length of the imaging system, X, Y, Z are co-ordinates of the object point and X_p, Y_p, Z_p are co-ordinates of the perspective centre at the time of imaging. Scale factor is denoted by s . M is the transformation matrix connecting the image and the object space. The matrix M is formed by multiplying a series of rotations connecting intermediate co-ordinate systems.

2.3 Relative Orientation of Multiview Images

The residual orientation error in the multiview mode of image acquisition is highly correlated for the overlapping images as

these images are acquired within a short interval of time from the same orbit. Thus, it is possible to perform analytic relative orientation of these images by considering the differential orientation error as unknown. This approach has two significant advantages; first, it is easy to identify the conjugate points in the overlapping images and secondly, if precise ground control points are not available, relatively oriented images can be used for computation of relative heights of the objects.

The developed approach to relatively orient the multiview images is based on coplanarity condition. It states that the perspective centres and the object point lie on the epipolar plane. Mathematically the coplanarity condition is expressed as

$$[\vec{a}_1, \vec{a}_2, \vec{b}] = 0 \quad (3)$$

where $[\cdot, \cdot, \cdot]$ represents scalar triple product and \vec{a}_1, \vec{a}_2 are the vectors joining object point and the perspective centre of the first and second images respectively, \vec{b} represents the vector connecting two perspective centres. Equation (3) is linearized with respect to differential roll, pitch and yaw values. Three pairs of conjugate points are sufficient to compute differential correction. The position of the perspective centres and the orientation information is primarily obtained from onboard/ground processed measurements supplied with the images.

2.4 Computation of Rational Polynomial Coefficients

Over the past decade, rational function models are being used as alternate to physical sensor models. This is primarily due to the fact that physical sensor models are complex; they need information about the camera geometry and good understanding of image acquisition process. On the other hand, rational function models are easy to implement and supported by major commercial satellite imagery providers. Moreover, using rational function model in place of physical sensor model makes the system truly sensor independent. However, it is important to quantify the results acquired with physical sensor model and the rational function model.

Rational polynomial coefficients are computed using terrain independent approach (Tao, 2002). The physical sensor model computes the orientation parameters as per the method explained in previous section. The image space co-ordinate is obtained for the given object space co-ordinate using the linearized form of equation (1) (Mahapatra et al, 2004). The image positions for a given set of uniformly spaced grid of object space co-ordinates are estimated using the physical sensor model. The set of object points and corresponding image positions are used to compute the rational polynomial coefficients. The derived set of rational polynomial coefficients are used for relating image and object space. Since the rational polynomial coefficients are used for further processing, it is possible to use commercially available stereo images obtained from satellites such as Geoeye-1, Worldview-1/2, and IKONOS for site model generation.

2.5 Image Matching Techniques

Digital image matching techniques are used for extraction of digital surface model and automatic identification of conjugate points. Digital image matching is considered as mathematically ill posed problem. This problem can be transformed to well posed one by imposing regularizing constraint. One possible technique is to reduce the domain of probable match by introducing geometric constraints. In building reconstruction problem line segments are automatically extracted and matched using geometric and photometric constraints (Baillard C, Park,

S, Schmid,C.). Some researchers proposed use of colour information for edge extraction and line segment stereo matching[Ok, A.O, Scholze, S.], our approach does not attempt to extract and match line segments for two reasons (1) only flat rooftops are modelled which means that same height is assigned to multiple points, thus, it is sufficient to match points on the edge of the building to compute height of the building. (2) Building boundaries are digitized, refined and connected on nadir image. These boundaries along with the height information analytically transferred to the ortho image as the relation between geometrically uncorrected image, epi-polar image and ortho image is established using the sensor model.

If two overlapping images are relatively oriented then the disparity between two conjugate points is due to the topographic relief. Based on the geometry either the set of points are matched in original stereo pair or the other image is re-sampled to form set of epi-polar images. In case of Cartosat-2, often the angle about roll axis is significant; this means that the disparity due to terrain variation will not be restricted to one dimension. Image matching strategy has to take into account the image acquisition process. The disparity map obtained in lower resolution is utilized to guide the matching in next step; In both cases geometric constraint is used to limit the search space. The dense DSM is generated using area based matching techniques performed on epi-polar images, while feature based matching matches the edges of buildings in raw as well as in epi-polar images using the geometric constraints. For both the image matching techniques, normalized cross correlation is used as a similarity measure. In both the matching approaches, template size is increased dynamically. The similarity measure of normalized cross correlation initiates with a small template and two thresholds, termed as noise and acceptance threshold. If the normalized cross correlation coefficient for the template is more than or equal to the acceptance correlation threshold, the point is checked for forward and reverse matching. It is accepted as a match point if the correlation coefficient is more than the acceptance threshold for both the directions. If the correlation coefficient is less than the acceptance threshold but higher than the noise threshold, the template size for reference and search space is increased and correlation coefficient is recomputed. The noise threshold is selected as 0.4 and acceptance threshold is selected as 0.9. Initial window size of the search image is 13 by 13 which is increased in steps of two pixels. The process is repeated for at least three different sized windows. This process is repeated for each level of image pyramid. The unmatched points are transferred to the next level using interpolation. The interpolation for unmatched point is not done for matching of original resolution images.

2.6 Computation of Normalized DSM

A Digital Terrain Model (DTM) is the elevation model of the landscape which does not include above ground objects. On the other hand, a Digital Surface Model (DSM) includes the objects with their heights above the ground as well as the topography. The man-made objects with different heights over the terrain can be detected by applying a threshold to Digital Surface Model. The DTM is estimated using mathematical morphology. The morphological operators help in bringing the background terrain from the DSM. The above ground objects are detected using the DSM and morph output. Two morphological operators, namely “opening” and “closing” are used iteratively. The size of the window depends upon the size of the building. The normalized DSM is generated by subtracting the DTM from the DSM. Segmentation and area filters detect the buildings of desired size. Finally the building outline can be

constructed using the neighbouring gradients. For the better definition of the building, the gradient information from the gray image is also used. The DTM provides the ground height which is used as an input for computing the building/object height.

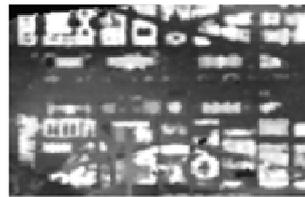


Fig. 6 Derived DSM



Fig. 7 Normalized DSM

2.7 Development of 2-D and 3-D Digitization Tool

Automatic detection of edges and grouping them to form a meaningful entity has been an area of research for a long time. Attempts to find a completely automatic and successful solution for these problems is an active area of research. A 2-D digitization option is used to manually digitize the building outlines in one image. These manually digitized boundaries are further refined using the Canny operator. In this case, the edges are found only in the neighbourhood of the manually digitized edge to ensure better localization. These edges are matched in other images using the geometrically constrained image matching procedure. In case the image matching procedure is not able to find corresponding edges in another image, digitization is done in 3-D viewing mode. User can move cursor in Z direction and place each vertex of the feature at any position in depth during digitization and fuse the cursors. Positions of cursors in left and right images are recorded. Options are available to draw points, lines and polygons. Fig. 2 shows digitized building boundaries for a portion of Washington image.

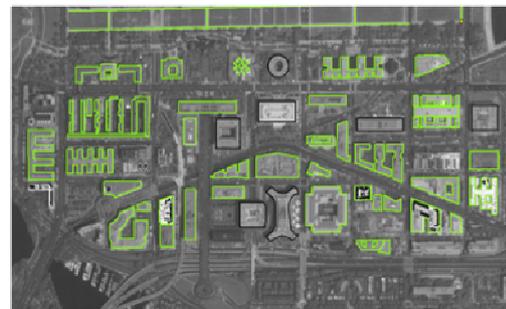


Fig. (2): Digitized building boundaries

The process improves edge localization and minimizes the effort of manual identification of edges precisely.

2.8 Refinement the Edges

Digitization of buildings is the major step in 3D site modelling, fully automatic methods for building extraction are not enough matured as a result we prefer semi automatic digitization method. The precision of edges is attained by refining the digitized edges to the nearest real edges which are lost in manual digitization. This is performed in the following steps,

1. Digitize manually in the proximity of the intended building.
2. Refine the edges that are manually digitized in step-1 by densifying and constructing a neighbourhood of each of these points and searching for real edges in canny edge image.

3. Join the disconnected edges guided by the digitized edges.

A simple case of digitizing (inner and outer edges of the circular building) a portion of image is shown in the figure 3(a). The orange boundaries show the manually digitized inner and outer edge of the building. The upper and lower circles in magenta of figure 3(b) are the annular region defined with suitable threshold; the edges shown in green are obtained through Canny operator [Canny, 1986]. The boundaries obtained through Canny operator provide better accuracy than manual digitization.

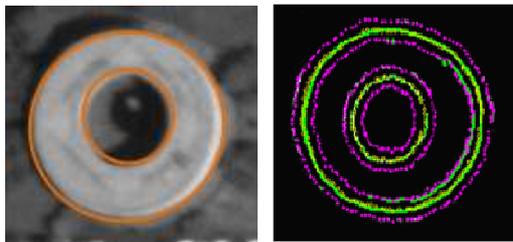


Fig. 3: (a) Digitized boundaries, (b) Refined boundary

2.9 Matching of Edges and DSM Generation

The DSM is obtained through the matching of the epipolar images at interval of four pixel units. Fig. 4 shows the obtained DSM. Fig. 5 shows the normalized DSM which is obtained from subtracting the derived DTM from the DSM. It is clear that the majority of the buildings can be detected in normalized DSM. It is intended to use the DSM as cue for building boundary extraction at later stage.

The positions of refined edges are known in near nadir image as shown in Fig. 6(a), and the corresponding points are estimated in the other image using the image to ground and ground to image transformation as shown in figure 6(b). Figure 6(c) displays matched points.

The position in one image is matched around the estimated position of the other image. The correlation threshold is chosen as 0.9. About 30 % of the points get matched. The height is computed for all these points which eventually represent the height of the building edge. The variations of calculated height at different points of the same roof top selected are of the order of 1m. The average height of all these points is assigned as height of the object assuming roof to be a plane surface. The building height with respect to ground is obtained by subtracting ground height derived from DTM.



Fig. 8(a) points on refined edges in nadir image

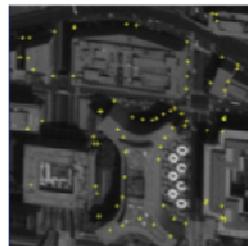


Fig. 8(b) Estimated points on image acquired with 26 deg view angle



Figure 8(c) Matched points on the edges of the image acquired with 26 deg view angle

2.10 Site Model Generation Process

The flow chart of site model generation process is depicted in figure 7. The basic inputs are at least two Cartosat-2 multiview images of the area of interest. The attitude and position information is available in ancillary data files. Using the physical sensor model, the relative orientation parameters are estimated. The rational polynomial coefficients are computed in terrain independent mode. The epipolar images are generated for these views. A dense DSM is used for generating normalized DSM and Digital Terrain Model. The edges of buildings are delineated by 2-D digitization and refinement procedures. The points on the edges are matched in another image. The remaining unmatched edges are manually digitized and refined in 3D viewing mode. The height is computed for the matched edge pairs. The ground level height is subtracted to get the building height. The height, delineated buildings and digital Terrain model are inputs for object modelling and visualization software.

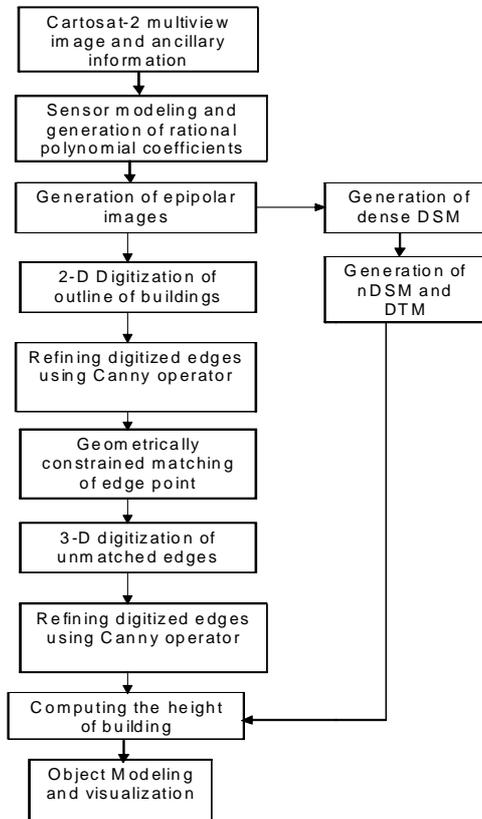


Fig 7: Block diagram of site model generation system.

3. RESULTS AND DISCUSSION

3.1 Results of Relative Orientation

Table 1 shows the results of relative orientation of multiview images. Fourteen conjugate points were identified on the overlapping images. Five points were used for computation of residual orientation parameters. The results are shown on remaining conjugate points. Starting with the image position in near nadir image, the image position of conjugate point is estimated in the second image. The estimated positions are compared against the actual positions, the difference between

the actual and estimated position is shown in line and pixel direction. The achieved average value is 0.02 and -0.001 pixels in line and pixel directions respectively. The standard deviation is 1.392 and 0.99 in line and pixel direction respectively.

Table 1 Results of Relative orientation of Multi-view images of Cartosat-2 images (Washington)

Actual line no	Actual pixel no	Estimated		Residual error (pixel units)		
		Line no	Pixel no	Along track	Across track	
18000.6	154.037	18000.6	153.967	0	0.07	
18164.3	509.96	18166.1	507.942	-1.8	2.018	
18234.6	507.42	18234.2	508.151	0.4	-0.731	
18216.2	875.47	18213.5	874.369	2.7	1.101	
18235.6	1043.02	18234.8	1043.31	0.8	-0.29	
18247.2	1134.07	18248.7	1134.07	-1.5	0	
18285.1	1429.53	18285	1429.88	0.1	-0.35	
18528.4	1966.17	18529.6	1967.36	-1.2	-1.19	
18518.4	2135.51	18517.7	2136.15	0.7	-0.64	
				std dev	1.392	0.99
				Average	0.022	-0.001

3.2 Comparison between rational polynomial coefficient and physical sensor model

The 'step n stare' mode of image acquisition is an asynchronous mode of imaging. It is observed that in this mode of image acquisition, the difference between the sensor model derived image position and Rational Polynomial Coefficient computed image position for same object point is of the order of 0.5 pixels.

Third order rational polynomial coefficients are fitted in terrain independent mode (Tao, 2002). In this mode the image position of a grid of equally spaced ground co-ordinates are computed using the physical sensor model. At least 200 points with height varying from minimum to maximum value in suitable step size are computed. These points are used to fit the rational polynomial coefficients. Image position of another set of ground points which do not have any point common with the set of points used for fitting the rational polynomial coefficients are computed using the physical sensor model and rational function model. The plot of the difference for near nadir image is shown in Fig. 8(a). The RMS error is 0.1244 and 0.090 pixel unit in line and pixel direction respectively for near nadir image.

The plot for image with a view angle of 26 degree is shown in Fig. 8(b). The blue line represents the residuals in line direction and the red line represents the residuals in pixel direction. The RMS error is 0.6910 and 0.5043 pixel unit in line and pixel direction respectively. These values are high compared to the residuals error between physical sensor derived and RPC derived positions for satellite like IRS-P6 where the mode of imaging is synchronous (Nagasubramaniam, 2007, Liang, 2006)

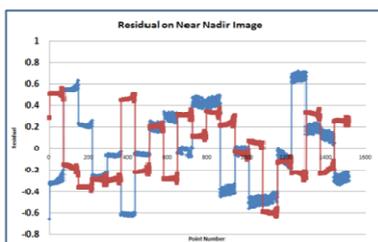


Fig. 8(b) Plot for nadir image

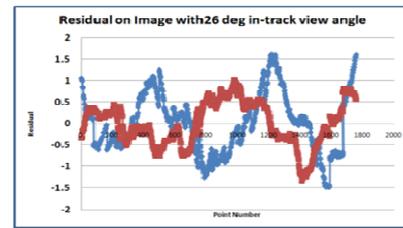


Fig. 8(b) Plot for image with 26 view angle

Fig. 8 Plot of difference between rational polynomial coefficient derived and physical sensor model derived image positions

3.3 Height Accuracies, Object Modelling and Visualization

The derived heights along with the building outlines are used and orthoimage are used as input for object modelling and visualization. Open source software Blender is used for creating triangular mesh and adding texture to the building. The texture is synthetic and symbolic; it may or may not represent the actual structure of the building. Fig. 9 shows a view of the generated site model. The height accuracies are evaluated with reference to LIDAR data available from Google Earth website. We observed that the average error is 0.8m and standard deviation is 1.8m.

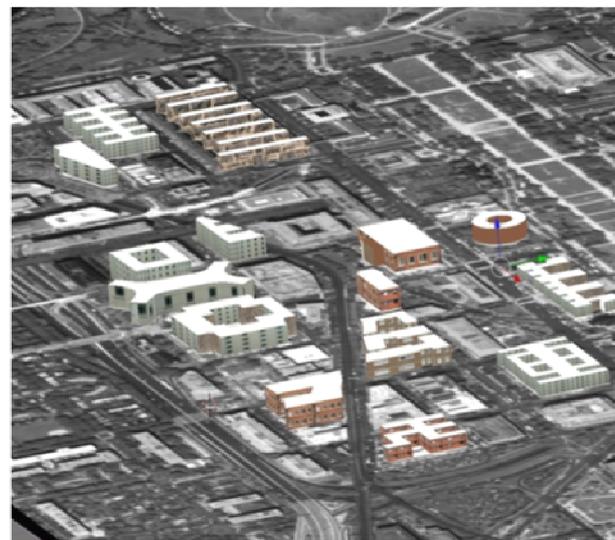


Fig. 9 A view of generated site model of portion of Washington city

4. CONCLUSION

The paper presents an overall schema and results for generation of site model from spaceborne multiview/stereo images. The results are shown for Cartosat-2 image. The relative orientation procedure obviates the need of precise ground control if relative measurements are suitable for the desired application. The system is designed in a generic way to accommodate images from other similar satellites if rational polynomial coefficients are available. At present the manual digitization process is starting point for capturing the outline of the buildings, this process too can be automated as the derived normalized DSM represents detection of buildings fairly well. Refinement of edges with Canny operator improves the edge localization. Geometrically constrained image matching procedure is

successful in finding conjugate points on edges. The system is primarily designed for spaceborne images, and the available resolution of Cartosat-2 and commercially available images permits extraction and representation of man-made structures. Detail roof reconstruction and representing complex roof structure may not be possible. A good photogrammetric framework provides the desired accuracy. Generated site models represent the area under consideration fairly well and can be used for various planning and other applications.

REFERENCES

- Baillard C., Schmid C., Zisserman A. and Fitzgibbon A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views, in ISPRS conference on automatic extraction of GIS objects from digital imagery, IAPRS volume 32, Part 3-2W5, pp. 69-80
- Blender 2.6.2, www.blender.org, last accessed on 20 April 2012
- Canny J F, 1986, Computational; approach to edge detection, IEEE transactions on Pattern Analysis and Machine Intelligence 8, pp 679-698..
- Gruen A, Xinhua Wang, 1999. Cyber City Modeler, a tool for interactive 3-D city model generation. In: German Photogrammetry Week, August, 1999, Stuttgart, Germany
- GoogleEarth, www.google.com/earth/explore/showcase/3dbuildings.html, last accessed on 10 January 2012
- Kraub Thomas, Manfred Lehner, Peter Reintarz, 2008. Generation of coarse 3-D models of urban areas from high resolution stereo satellite images. In: International archives of the photogrammetry, remote sensing and spatial information sciences, vol XXXVII, part B1, pp 1091-1098, Beijing
- Liang-Chien Chen, Tee-Ann TEO, Chien-Liang Liu, May 2006, The Geometrical Comparisions of RSM and RFM for Formosat -2 Satellite Images. *Photogrammetric Engineering and Remote Sensing Vol.72 No:5*, 573-579.
- Mahapatra Archana, R Ramchandran, and R Krishnan, 2004. Modeling the Uncertainty in Orientation of IRS-1C/1D with A rigorous Photogrammetric Model. PE & RS, 70(8), pp 939-946
- Nagasubramanian V. , Radhadevi P.V. , Ramachandran R. , and Krishnan R. 2007. Rational function model for sensor orientation of IRS-P6 Liss-4 imagery. The Photogrammetric Record, 22(120), pp 309-320.
- Ok, A.O.; Wegner, J.D.; Heipke, C.; Rottensteiner, F.; Soergel, U.; Toprak, V. (2010): A new straight line reconstruction methodology from multi-spectral stereo aerial images, In: Proceedings of Photogrammetric Computer Vision and Image Analysis Conference, IntArchPhRS vol. XXXVIII, part 3A, Paris, 2010, pp. 25-30
- Park, S., Lee, K., and Lee, S., 2000. A Line Feature Matching Technique Based on an Eigenvector Approach, Computer Vision and Image Understanding, 77, pp. 263-283.
- Scholze, S., Moons, T., Ade, F., Van Gool, L., 2000. Exploiting Color for Edge Extraction and Line Segment Stereo Matching. In: IAPRS, 815-822.
- Schmid, C. and Zisserman, A., 1997. Automatic Line Matching Across Views. In: Proceedings of CVPR, pp. 666-671.
- Tao C.V. and HU, Y., July 2002. 3-D Reconstruction methods based on the Rational Function Model, *Photogrammetric Engineering and Remote Sensing Vol. 68, No: 7*, 705-714.
- Ulm K and Daniela Poli. 3D city with Cybercity modeller. In: 1st EARSeL Workshop of the SIG Urban Remote Sensing Humboldt-Universität zu Berlin, 2-3 March 2006.
- Vosselman G, Dijkman D, 2001. 3D building model reconstruction from point clouds and ground plans. In: International Archives of Photogrammetry and Remote Sensing, Vol XXXIV-2/W4 Annapolis, MD