Disparity map generation from satellite stereo pair images and estimating height information from it using ANN and SVR

Arati Paul, Krishula Sinha

Abstract— **Stereo vision is generally used in depth information generation. The classical approach involves imaging geometry for generating DEM or DSM from a pair of stereo images. In the present study a software tool has been developed to extract relative depth information of earth features, viz. buildings, in terms of disparity map, from a pair of images with different viewing angles. Any images of same area acquired in different perspective can be used to generate disparity map using this tool without the knowledge of their imaging parameter. The disparity values are subsequently compared with the LiDAR generated DSM values and a strong correlation has been found between them. Using machine learning algorithm viz. ANN and SVR, heights of unknown ground objects have been predicted from disparity values with an appreciable accuracy.**

Keywords— **disparity, height, ANN, SVR, stereo**

I. INTRODUCTION

Generation of elevation information from satellite stereo pair images is widely discussed topic in digital photogrammetry and quite a large no of appreciable works have been reported regarding DEM or DSM generation. The general approach for automatic DEM generation involves mathematical models to georeference the stereo images, which can be categorized into Rigorous model [3] and Generic model [4, 7, 9].Generic model is used in the absence of attitude parameters. Subsequently, epipolar images are been generated to reduce the time of finding corresponding points in image matching [10]. Finally the DEM is generated after accuracy assessment with check points.

Very high resolution (VHR) satellite like Ikonos, QuickBird, GeoEye or WorldView I and II, does not always "look" vertically.

Thus images of different perspective of same area may be available from different sources viz. Google earth. Sometimes they may not be perfect stereo pair where perspective geometry of the imaging system as well as RPCs is known. This kind of image pair may not directly be used to generate elevation information using existing classical method. Disparity map, generated from a pair of images acquired in different perspective also caries relative depth information and it can be generated by image comparison without the knowledge of any other imaging parameter. Some of the disparity map generation methods can be found in [6] along with their performance. Most of these automatic disparity map generation algorithms work well for real life images and it is difficult to achieve desired result for satellite images.

In the present study an interactive disparity map generation tool has been developed to compute disparity of any earth feature between two images. Subsequently disparities of ground objects (viz. building) have been computed for the scene and compared with LiDAR generated DSM. Further the study has been extended to estimate elevation information from disparity values using machine learning algorithms viz. ANN (Artificial Neural Network) and SVR (Support Vector Regression) and a very good correlation has been observed between the estimated and actual elevation value. This tool enables estimating height information by image comparison without using imaging geometry of actual stereo pair and is also capable of generating height information from images which are available freely from sources like google earth. Hence it can be considered as a low cost solution for generating a height map of a particular area. The detail methodology and results are discussed in subsequent sections.

II. OVERVIEW OF THE PROPOSED METHOD

Tall objects yields more positional shift in stereo vision. This positional shift can be captured from a stereo pair images in a disparity map, which in turn can revel relative height information of objects in the image. In present study a disparity map

generation tool has been developed which computes the positional difference between two conjugate objects in left and right image interactively and generates disparity map without using any imaging geometry or other associated information. This method is useful when metadata of image is not known. Some examples of such images from Google earth are shown in figure1 where two different date images show different perspective and can be used to generate relative height information. In present study Worldview-1 Stereo pair data from ISPRS Matching Benchmark datasets has been used, and the output has been compared and tested with available data. Further, the study has been extended to estimate height information from disparity map using machine learning algorithms viz. ANN, SVR where height information for some location has been considered to be known from other source. For experiment purpose LiDAR generated DSM has been considered as a reference for height information. As depth is inversely proportional to disparity, these learning models show good accuracy in estimating height information. The schematic diagram of the methodology has been shown in figure2.
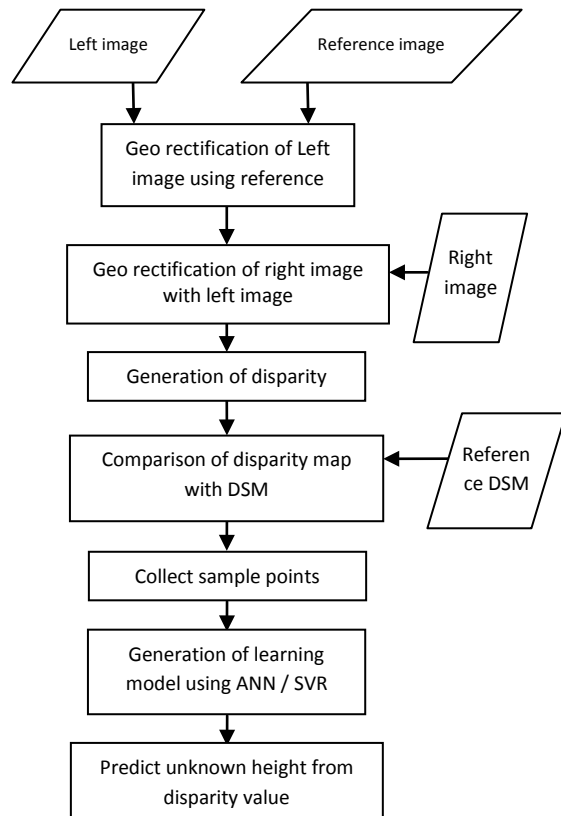

(a)

(b)

Figure 1: Goole earth image of18°32′51″ E, 73°54′41″ N location (a) 29/1/2013 image, (b) 2/11/2013image


Figure 2: Schematic diagram of the proposed method

III. DATA USED AND PREPROCESSING

The proposed methodology has been demonstrated using three different data sets. Following sections describe the data used and pre-processing of data briefly.

A. Reference data

Landsat ETM+ band 8 data with path- row as 198-031has been used in the study for georeferencing the left image of the stereo pair (Figure 3). The data is projected in UTM WGS84, Zone 31 North.
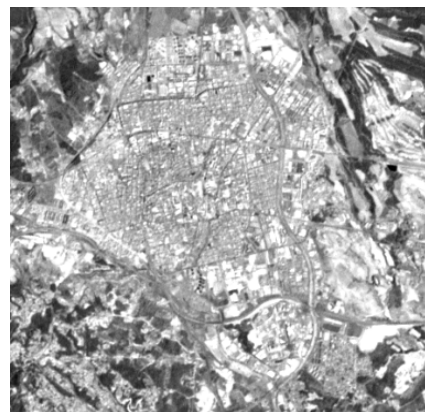
Figure 3: Landsat ETM+ band 8

B. Stereo data

To demonstrate the methodology the ISPRS Matching Benchmark datasets [8] are used. The details about the data can be found at (ISPRS, 2010).Worldview-1 Stereo pair acquired on the 29th August 2008 over Terrassa, Spain (417400E 4597300N) covering landuse type as City, industrial and residential is used (figure 4). One image was acquired with an off nadir angle of 4° (referred as left image in this article) and a spatial resolution of 0.5 m, the other with an off nadir angle of 33° and a spatial resolution of 0.66 m (referred as right image).Although the method does not require a perfect stereo pair image with associated metadata, the Worldview-1 has been chosen for its availability. The stereo pair has been used to generate disparity map.


(a)

(b)

Figure 4: Input stereo images (a) left image, (b) right image

C. Test data

For testing the disparity values 3D point cloud acquired by airborne laser scanning with a density of approximately 0.5 points per square meter has been used which was provided by the same source that of the stereo data. The LIDAR data for the Terrassa test area was acquired on 26th and 27th November 2007. DSM has been generated using the LiDAR data (figure 5).
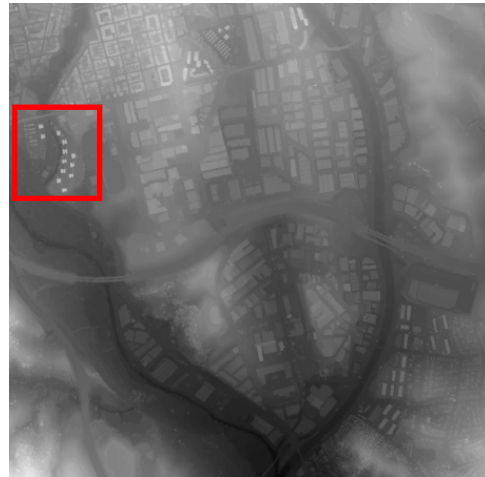
Figure 5: LiDAR generated DSM

D. Data pre-processing

The left image of the stereo pair has been georeferenced using Landsat image as reference. 2nd order polynomial has been used as geometric model and the output has been resample using cubic convolution. The right image is then registered with the left image using the same model parameter. In this step the main objective was to match the ground features viz. roads in both the images. However buildings which change its position with different perspective were allowed to maintain its default position in right image after registration. The geo referenced left and right images are used for disparity map generation. Subsequently DSM also has been generated from LiDAR (.las) data taking only the first pulse return for comparing the disparity values.

IV. DISPARITY MAP GENERATION

Depth information from two images of same scene can be derived in an intuitive way. Figure 6 demonstrates a stereo imaging system and it contains equivalent triangles.
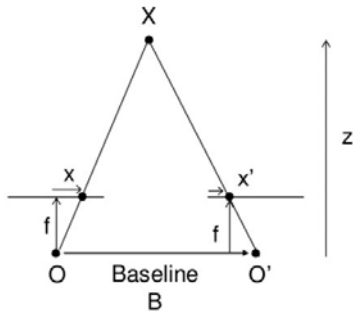
Figure 6: Stereo imaging system

Where x and x' are the distance between points in image plane for left and right camera. B is the distance between two cameras and f is the focal length of camera. Z is depth of the object X. Writing their equivalent equations will yield equation (1)

$$Disparity = x - x' = \frac{Bf}{z} \quad (1)$$

$$Depth \propto \frac{1}{Disparity} \quad (2)$$

Hence the depth of a point in a scene is inversely proportional to the disparity (equation 2). In absence of camera parameter disparity map gives relative depth information.

A. Review of selected alternate methods

The issue of disparity map generation from stereo pair images has been studied several times. Correlation based similarity measure is a commonly used approach for disparity map generation. Some of the well-known correlation based methods are sum of absolute differences (SAD), sum of squared deviations (SSD) and normalized cross correlation (NCC) [1]. These algorithms are applied on image pair (figure 7a & b) and the disparity map generated from SAD, SSD and NCC are shown in figure 7c, d and e respectively.
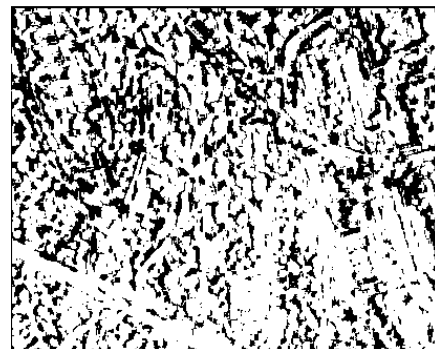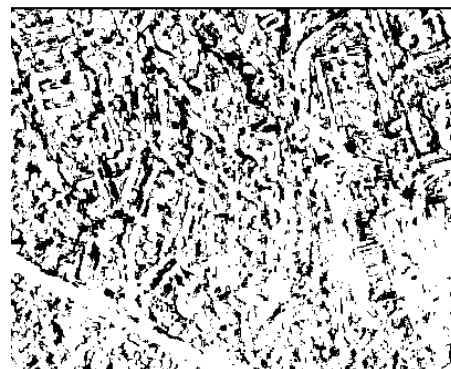

(a)

(b)
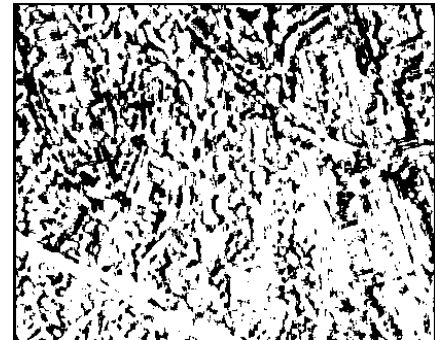
(c)

(d)

(e)

Figure 7: Subset of (a) Left Image, (b) Right Image and output of (c) SAD (d) SSD (e) NCC

B. Interactive disparity map generation tool

In present study an interactive disparity map generation tool has been developed as the automated methods as discussed were yielding unsatisfactory and noisy result. The tool takes the stereo images as input and displays it. When a conjugate object pair is drawn over the left and the right image, their centroid positions are calculated. The Euclidian distance between two centroids gives the disparity value for that object pair. Figure 8 shows the GUI of the tool along with object conjugates and other functionalities. Disparity map is generated using this tool for the study area and a subset is shown in figure 9, where different disparity values are represented in different grey sheds. In the figure 9 it is seen that different disparity values are generated for different height objects (or building) without any noise.
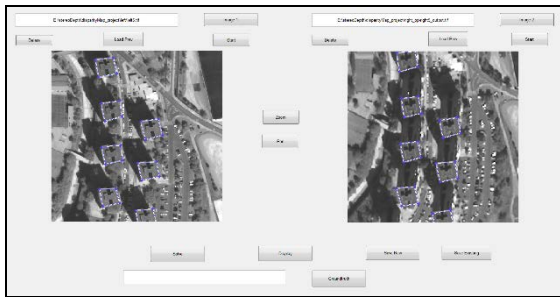
**Figure 8:** Disparity map generation tool
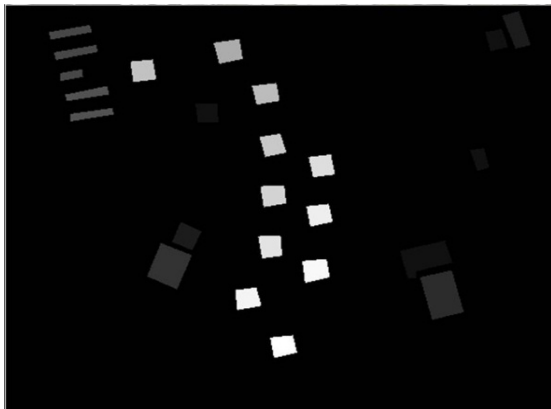
Figure 9: Subset of Disparity map

V. MACHINE LEARNING FOR GENERATION DEPTH INFORMATION FROM DISPARITY

Disparity map gives relative depth information which in turn may be used to generate height information through machine learning algorithms if actual ground truth is available for some location. In the present study the LIDAR generated DSM has been used as a ground truth image. Scatter plot of computed disparity values and its corresponding ground truth values for few observations is shown in figure 10. A strong correlation is clearly visible between these two parameters that indicate the possibility of exploring machine learning capability to establish a model between disparity and ground truth and predict the ground truth value for a given disparity value whose actual height is unknown.

Two different machine learning process viz. artificial neural network (ANN) and support vector machine based regression (SVR) have been applied on the data set to predict the height values for given disparity values. The performance of the model has been estimated using R^2 (square correlation coefficient).
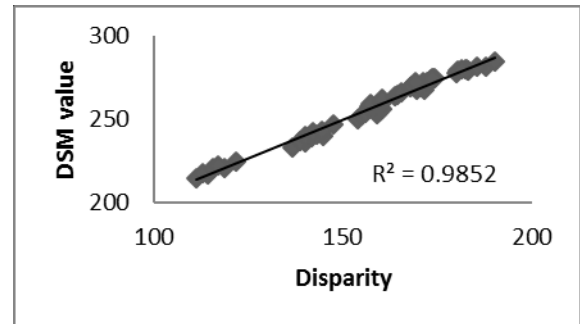
Figure 10: Disparity vs actual Height

A. ANN based prediction

The total 58 observation has been divided randomly into Training (40%) and Testing (60%). Number of hidden neurons was 5. The Marquardt-Levenberg algorithm [5] was found to be best suited training algorithm and was used to train the network. The result of observed and ANN estimated height are plotted in figure 11. A very high correlation (R^2 value 0.99) between actual and estimated values has been observed.
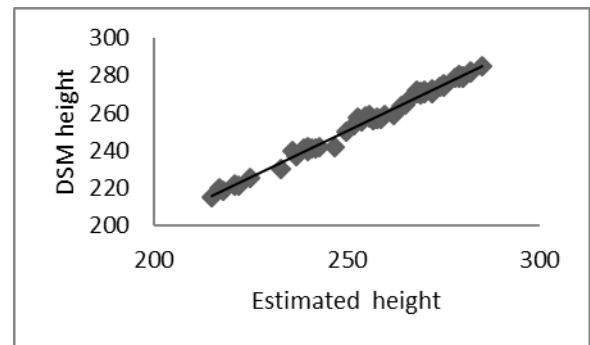
Figure 11: DSM vs. ANN predicted Height

B. SVR based prediction

Support vector machines (SVMs) are supervised statistical learning models, characterized by usage of kernels to identify patterns for classification and regression analysis [2]. In the present study the SVR with capacity (C) value 5.000 and epsilon value 0.100 has been chosen. The kernel type is Radial Basis Function (RBF) with gamma value 1.000. The total sample (58no.s) has been divided into training (43) and testing (15). The R^2 value of measured and SVR estimated height has come as 0.99 which also can be considered as very high (figure 12).
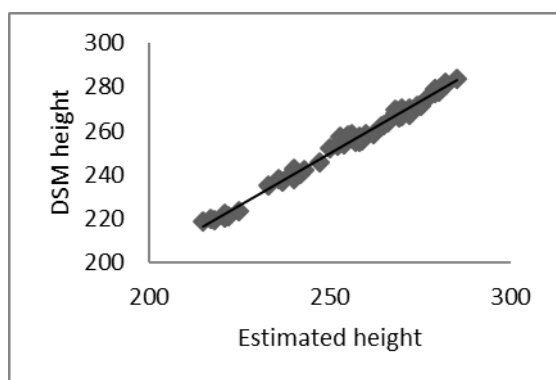
Figure 12: DSM vs. SVR predicted Height

VI. SUMMARY AND CONCLUSION

The disparity map generation tool calculates the Euclidian distance between the conjugate object pair which has been generated by the parallax shift between two images. The methodology discussed here is capable of generating disparity map from a pair of images of different perspective without using any other information. From the above mentioned results it can be concluded that height information can be predicted from disparity values with a good accuracy using any of the above mentioned learning algorithms if training data is available for a particular geographical area. This is an indirect approach of getting height information without using photogrammetry. Ground objects viz. roads, fields, water bodies etc. of left and right images need to be registered perfectly for accurate disparity estimation. However, the proposed methodology is simple and good accuracy between disparity and ground truth as well as between predicted and actual ground truth shows the success of the method in generating height information from any two images of same area with different perspective. This kind of height information can be used in city modelling.

References

[1] Ahuja, S. 2009. Correlation based similarity measures. www.siddhantahuja.wordpress.com /2009/05/20/correlation-based-similarity-measure-normalized-cross-correlation-ncc.
[2] Cortes, C. and Vapnik, V. 1995.Support-vector networks. *Machine Learning* 20 (3): 273
ISPRS, http://www.commission1.isprs.org/wg4/
[3] Kornus, W., Alamús, R., Ruiz, A., Talaya, J., 2006. DEM generation from SPOT-5 3-fold along track stereoscopic imagery using autocalibration. *ISPRS Journal of Photogrammetry & Remote Sensing.* 60:147–159.
[4] Krauß, T., REINARTZ, P., LEHNER, M., SCHROEDER, M., STILLA, U., 2005.DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. *ISPRS Hannover Workshop 2005 on "High-Resolution Earth Imaging for Geospatial Information",* Hannover, Germany, 17–20 May. On CD-ROM.
[5] Lubna, B. Md, Hamdan Md. A., Abdelhafez E A. And Shaheen W. 2013, Hourly Solar Radiation Prediction Based on Nonlinear Autoregressive Exogenous (Narx) Neural Network. Jordan Journal of Mechanical and Industrial Engineering, 7(1):11-18, ISSN 1995-6665.
[6] Middlebury, http://vision.middlebury. edu/ ~schar / stereo/web/ results. php
[7] Madani, M. 1999, Real-Time Sensor-Independent Positioning by Rational Functions, *Proceedings of ISPRS Workshop on Direct Versus Indirect Methods of Sensor Orientation, Barcelona, Spain, November 25-26,* 64-75
[8] Reinartz, P., D'angelo, P., Krauß, T., Poli, D., Jacobsen, K. And Buyuksalih, G. 2010. Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data. *ISPRS.*
[9] Sadeghian, S., Zoej, M. J. V., Delavarl, M. R., Abootalebiz. A. 2001.Precision rectification of high resolution satellite imagery without ephemeris data. *JAG.* 3(4).
[10] Zhang, L., Gruen, A. 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry & Remote Sensing.* 60: 195–211.