# 3D Reconstruction from Stereoscopic Satellite Images

Tanmay Gatle1, Manasi Ladkat2, Anurag Kolhe3, Sunil B. Mane4, Rohini Sarode5

*1,2,3Department of Computer Engineering and IT, College of Engineering Pune, Pune, Maharashtra, India*
*4,5Assistant Professor, Department of Computer Engineering and IT,*
*College of Engineering Pune, Pune, Maharashtra, India*

**Abstract- We develop a Web Tool for 3D Reconstruction of Satellite Images. We take input as several images of a single site and generate a resulting 3D model of the same site. This method uses local affine camera approximation for processing, and thus generates 3D models of small areas only. Usually, if the images of a single site are taken on different dates or in different seasons, then there are many changes in vegetation. Hence, the resulting 3D model will not be as accurate as compared to the model we would have generated by inputting images on a single date. However, our method generates 3D models with significant accuracy even if the images are taken on different dates or in different seasons. For this, we propose a Pair Selection criterion which solves this problem to a greater extent. We select a good stereo pair of images and generate a DEM from that pair. Similarly, some few good pairs are found by pair selection and resulting point clouds are aligned and fused together. Hence, it shows that inputting a large number of images of a single site and refining the 3D model by pair selection gives a 3D model which is as accurate as a single same date stereo pair.**
**Keywords – Stereoscopic, Satellite, Rectification, Triangulation, Point Cloud**

## I. INTRODUCTION

The availability of high-resolution satellite images has increased in recent years. Space borne sensors like QuickBird and IKONOS provide not only for high resolution and multi-spectral data but also for the capability of stereo mapping. The use of stereoscopic imagery and other sensor models such as rational function models can be used for 3D reconstruction of satellite imagery with accurate depth. The most important applications of 3D reconstruction are in Defense and Smart City modeling and can be used to keep an accurate account of the suburban and rural areas. These 3D models can also be used for disaster simulations like earthquake and flooding, security evaluation, watch over illegal constructions, pollution monitoring, and environmental planning. Other areas of application are change detection, cartography and large-scale measurements such as assessment after natural disasters and forest evolution.

The images obtained from satellites have led to more accurate maps of the world being generated. The use of more advanced technology has also increased the accuracy and quality of the pictures obtained. Due to a mass of satellites available, many sites are even surveyed several times a year. Most of the images were multi-date images. Although not much change is observed, a closer look at the pictures shows a significant difference in vegetation and changes in the sky as per the climate. The goal of this project is to present an application which takes as input a variety of satellite images belonging to a particular site and to compute the best possible 3D model with the least computational cost.

### 1.1 Coordinate Systems –

In this project, we need to define geographical locations in terms of its relative coordinates corresponding to the earth, i.e., the geographical coordinates. For this, we use a coordinate reference system (CRS) where the coordinates are defined in terms of latitude, longitude and altitude.
To define the location of a particular geographical point on the image we use the term pixel coordinates. It defines the location of a pixel in the image uniquely. A digital image consists of row and columns of pixels. The pixel coordinates are a multidimensional array consisting of the row number and column number of the position of the pixel.

### 1.2 Dataset –

The dataset provided by IARPA Multi-view Stereo 3D Mapping Challenge is considered for the project. It consists of 50 DigitalGlobe WorldView-3 panchromatic images from its sample set of publicly available images. The images were acquired over a period of 14 months and have a 30 cm nadir resolution. The images are affected by the climate changes; the clarity of the images from the winter months is considerably lower. While using the dataset we consider an area described by geographic coordinates to be the area of interest (AOI). We want to focus on this part of the image for 3D reconstruction. The AOI in each of the images of the dataset is considered for further computations.

*1.3 RPC Camera Model –*

Each satellite image is associated with some metadata. This metadata consists of the exact acquisition date, the direction of the sun, the Rational Polynomial Coefficients (RPC) camera model, etc. The images acquired from the satellites are of a particular area on the surface of the earth. Thus, the user must be able to calculate the geographical coordinates corresponding to any point of the image. Thus, the pixel coordinates must be converted into geographical coordinates and back.

This function is provided by the Rational Polynomial Model (RPC). The RPC can be calculated by understanding the orientation parameters of the cameras. Instead, each image is provided with an RPC model consisting of localization and projection functions. The localization function takes an image point coordinates as input and returns the geographical coordinates of that point. The projection function takes the geographical coordinates as input and returns the pixel coordinates of the image. The RPC coordinates are quite accurate but they are subject to measurement errors. This error in the translation can lead to geopositioning errors of the triangulated points. The errors can be of the order of tens of pixels in the image domain.

## II. LITERATURE REVIEW

*2.1 Overview –*

This is a Web Tool for generating 3D models of a particular place. The user will provide several Satellite Images of a particular site of which the 3D model is expected and our tool will try to generate the best 3D model possible. Satellite Images contain huge metadata such as latitude, longitude, RPC functions, etc. which will help us in this process.

*2.2 Existing Technologies and Work in Domain –*

Due to the need to monitor changes on the surface of the earth, the number of earth observation satellite has increased very much in past few years. Hence, there are a lot of images of the same site available especially sites in urban areas. New technologies result in high-resolution satellite images which can be useful for identification of surface objects.

Currently, there are two methods by which the problem of 3D reconstruction can be solved. These methods are –

True Multiview - In this method, a 3D model from several images is generated by doing the process of triangulation for all images simultaneously.

Multiview Stereo – Unlike True Multiview, instead of considering all images simultaneously, this method uses binocular stereo to select several image pairs independently and hence fuse the resulting models. It is observed that this method works better than the previous one.

For the formation of 3D models, Stereo Matching is the most essential step. In brief, it's a step which matches the pixel from one image to the corresponding pixel in another image. It includes two steps - Epipolar Rectification and Triangulation.

Epipolar Rectification is a step which causes the resampling of images such that depth variations cause apparent motion in the horizontal direction only. There is a one-to-one correspondence be- tween epipolar lines in Pinhole camera. However, Satellite cameras are pushbroom and hence there is no one-to-one correspondence between them. Hence, this step is necessary to improve speed, accuracy, and robustness.

Triangulation is the process in which a 3D point is generated from a pair of images by identifying a point correspondence between them. RPC functions known as projection and localization helps in this process. These functions are provided in the metadata of the satellite images. After doing this process Stereo Matching is completed, and we get a DEM.

A module in python named OpenCV is used for 3D reconstruction of an object. Initially, the camera calibration is done by following the instructions given in the OpenCV tutorial. After that several images of the objects are provided. OpenCV will create a depth map by performing stereo matching. Quality of the final 3D object depends on the quality of the images provided (i.e. the lighting effects, angle of the photo, etc.) and also on the camera calibration. A module named open3d can be used to display the .ply files which are generated by OpenCV. Matlab is further used for alignment and fusion of the point clouds.

*2.3 Problems in Existing Technologies –*

As mentioned earlier that there are two techniques to generate 3D models, i.e True Multiview and Multiview Stereo. This, however, can be an issue when dealing with multi-date images whose area of interest (AOI) is small.

Hence, the problems faced are –

Input images are from different satellite.

Input images belong to different dates.

Input images are taken in different seasons which may result in vegetation changes.
There may be radiometric changes.
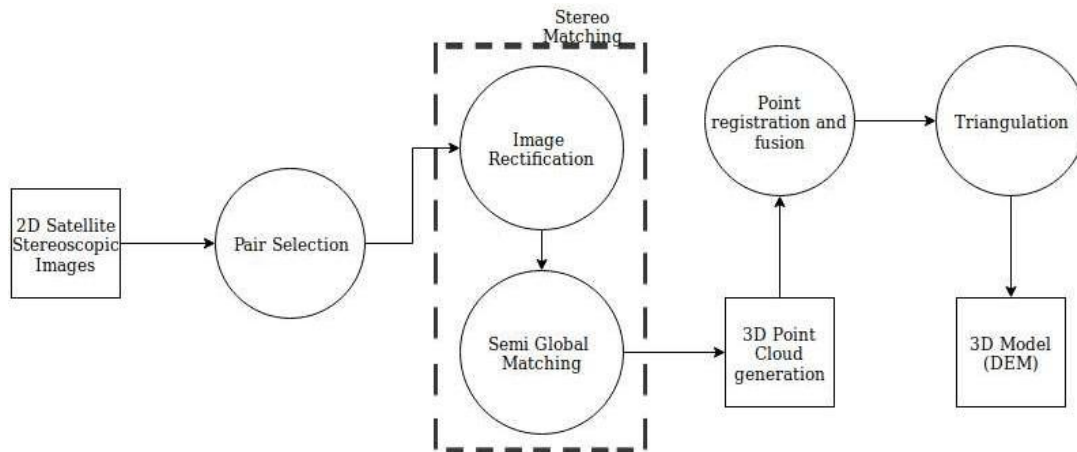Satellites use a Pushbroom camera instead of Affine camera.



Figure 1: 3D Reconstruction steps.

## III. PROPOSED ALGORITHM

As stated earlier, the Multiview stereo approach gives better results. All of the methods for 3D reconstruction rely on bundle adjustment for correcting 3D inaccuracies. This can be an issue when images are multi-date and AOI is small.

So we will propose a method like Multiview stereo but without bundle adjustment i.e. we will do stereo matching using independent pairs and then fusing all images together without doing any prior bundle adjustment. Hence we try to show that if we select best image pairs to compute our 3D model then the result obtained from multi-date image pairs can be as accurate as single-date stereo pairs.

Multiview stereo takes the median of all the independent pairs to give the final result. Instead of this, we propose to fuse and align only selected pair of images. So following are the steps involved:

*3.1 Pair Selection –*
The qualities of the 3D model obtained from different pair of images vary widely. Some result in very accurate 3D models while others result in very bad 3D models. Factors such as seasonal changes, shadows, radiometric changes, etc. affect the quality of the resulting 3D model. So, we need a strategy to sort the pair of images based on some sort of criterion. Pair of images which qualify this threshold will only be considered for generating DEM's.

In order to do Pair Selection, we need an area of interest to proceed. So, first, we need to crop the AOI from the given image. First, we create an array consisting of coordinates of AOI. We then calculate the coordinates of the center of the AOI by taking a mean over all the coordinates of the AOI.

Now, we need to convert these coordinates into pixels so we can crop a part of the image. We are using a GeoTIFF image. Its metadata contains RPC functions along with other information. A projection function is provided in its metadata which converts a 3D point defined by latitude, longitude and height into a 2D image pixel. In python's 'utils' module, a function is provided to extract RPC functions from a given GeoTIFF image.

Now, we can use the projection function in order to get the pixels corresponding to the AOI coordinates defined earlier. But the problem here is that projection function along with latitude and longitude also requires height i.e. a z-coordinate. We may assume that z is 0 or alternatively get z from an external Digital Elevation Model such as SRTM. The SRTM altitude at a given latitude and longitude can be obtained using the srtm4 module. We get the value of z by passing the latitude and longitude values of the center to the SRTM function present in the SRTM module. It then gives the value of z which can be used by projection function.

Hence, now parameters to the projection function are - an array containing latitudes of AOI, an array containing longitudes of AOI, z-value obtained using the srtm4 module. Projection function returns two arrays each containing the x and y pixel values corresponding to the AOI coordinates. So finally we can crop the AOI out of the given image.

Now, consider the census cost for stereo vision as it is more robust. We compute the census cost for two image pairs. This returns a cost volume matrix which consists of different disparity values for each point. In order to

proceed, we need to get the value of a point for a single minimum and accurate disparity. Hence, we do processing on the 3D array such that the minimum disparity and minimum cost is selected for each point. This will give 2D array with each point and its corresponding disparity value and then we can compute the variance of the given values.

We judged the reliability of the image pair based on the total variance of the 2D array of disparities. Now, if the variance value is very low then it's too uniform and close and not a good approximation for the given images as shown in figure 2. And if the variance value is too high then the resulting point cloud is too distorted as mentioned in figure 3. So according to our observation for our dataset image pairs with variance in range 62-67 produced optimal results. One of such results is shown in figure 4. Image pairs outside these values produced distorted and poor results. Hence, we select the pairs having their variance in the stated range for further processing.
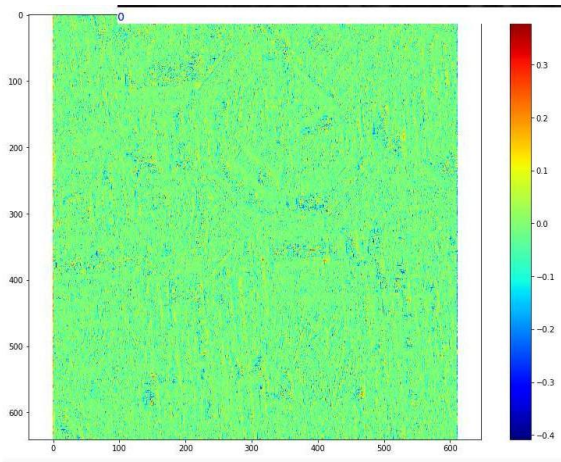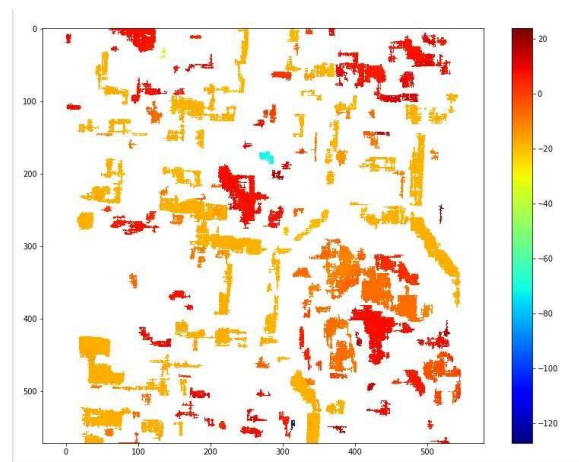


Figure 2: DEM of low variance
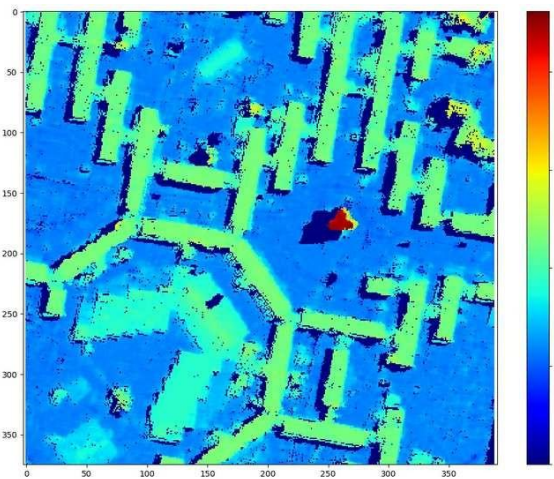


Figure 3: DEM of high variance



Figure 4: DEM of optimum variance

*3.2 Rectification –*

In order to ensure smooth stereo matching, we have to first align the two satellite images such that the epipolar lines of both images coincide horizontally. The corresponding points in both the images can then simply be searched for along these lines. This reduces the complexity of the searching problem. The rectification process involves finding two matrices which denote the translation, rotation and zooming of the two cameras such that the images obtained are rectified.

Let P be the projection function of the camera. The Taylor approximation of P around a point X0 can be written as
P(X) = P(X0) + P(X0)X + T
Where T = P(X0) - P(X0)X0. This can be written as a linear operation in the following way
P(X) = AX

where A is the affine i.e. linear approximation of the P at point X0. This A matrix basically resembles the projection function of a point in 3D space to image coordinates. In this way, both the images may or may not have the same RPCs and in turn affine approximation matrices. So first we find the A matrix for each image denoted as S1 and S2 from each of the images corresponding RPC values. Then these matrices should be modified in such a manner that they define the rotation, translation and zoom of the cameras (or images) in order for the resultant images to be rectified i.e. their epipolar lines are collinear. After computing the transformation matrices, the width and height of the AOI along with the corresponding transformation matrix is applied to each image to get the rectified image.

Along with rectification, we have to compute some parameters which are to be fed to the stereo matching algorithm. Firstly, we compute the SIFT key point matches of the two original images using the OpenCV tool. These are the corresponding points of the two images, on which we apply the transformation matrices to get the final rectified points. By subtracting each of the corresponding points we get the disparity range which is to be considered during stereo matching. This disparity range is then fed into the stereo matching algorithm.

*3.3 Stereo Matching –*

Stereo matching is the process of inferring depth information from two stereo images i.e. horizontally displaced images. Once the images are rectified, the depth of the objects can be computed by the difference in their horizontal coordinates. The depth of a point is inversely proportional to this difference. This information is used to build a disparity map.

The stereo matching algorithm used is Semi Global Matching (SGM) developed by Heiko Hirschmueller. It uses a dynamic programming approach where we slide a window on both images along the epipolar line and store the cost in a table or matrix. For this purpose, the images should be stereo rectified as then the corresponding pixels will lie on the same line. Doing this will make the algorithm computationally less costly.

Stereo vision cost can be computed by usually squared difference(sd) but it's not quite robust to illumination changes. Some studies identify Census as the most robust cost for stereo vision. Here a bit vector stores specific values depending on whether the intensity values of neighboring pixels are high or low as compared to the center pixel. This is also called a census transform. Pixel-wise matching is done by computing the Hamming distance of bit vectors of corresponding pixels. Census is well suited of local radiometric changes as it considers computation on local neighborhood only. SGM uses hamming distances pairwise and not the sum of hamming distances over another window.

The cost is computed for each pixel in the window with each disparity in the disparity range i.e. from dmin to dmax for different orientations or scanlines. Then we sum up the cost values of each window along each orientation for each disparity value in the disparity range. We assign the disparity value for each window which has the minimum aggregated cost. The stereo matching cost used is Hamming Distance between Census Transforms due to its robustness to radiometric changes.

However, the disparity map obtained may still be imperfect and distorted. In order to check for incorrect matches, we first apply a left-right filter. There are also some regions which are visible in one image and not visible in the other due to the change in the position of the camera. These regions lead to the formation of speckles which are removed using the speckle filter.

*3.4 Triangulation –*

Image correspondences obtained during the previous step are used to extract 3D points. These 3D points are compiled into a single point cloud and projected using Open3D to obtain a 3D model. For this, a disparity map between the chosen image pairs is first generated. This shows the pixel difference between the pair of stereo images. Each disparity in the disparity map is then used to calculate the 3D projection of a particular point. We use a function which is given a disparity map and RPCs of the pair of images as input. All the valid disparities are aligned into a single array.

For each disparity, the point where the disparity occurs is determined and the pixel coordinates are calculated using the RPC model. The affine coordinates as well the pixel coordinates are known. The only unknown is the height. This is solved by defining the problem as two linear equations in one unknown. We use the localization function as we know the affine and pixel coordinates and the height is the unknown. We have two such images corresponding to each of the stereo images. This system can be interpreted as the intersection of two lines in 3D space. The longitude, latitude and height obtained are then used to find the UTM coordinates for that point. All the UTM coordinates are added to an output list which forms the point cloud of the given images. This point cloud is then used to generate a Digital Elevation Model (DEM) which is displayed as the result.
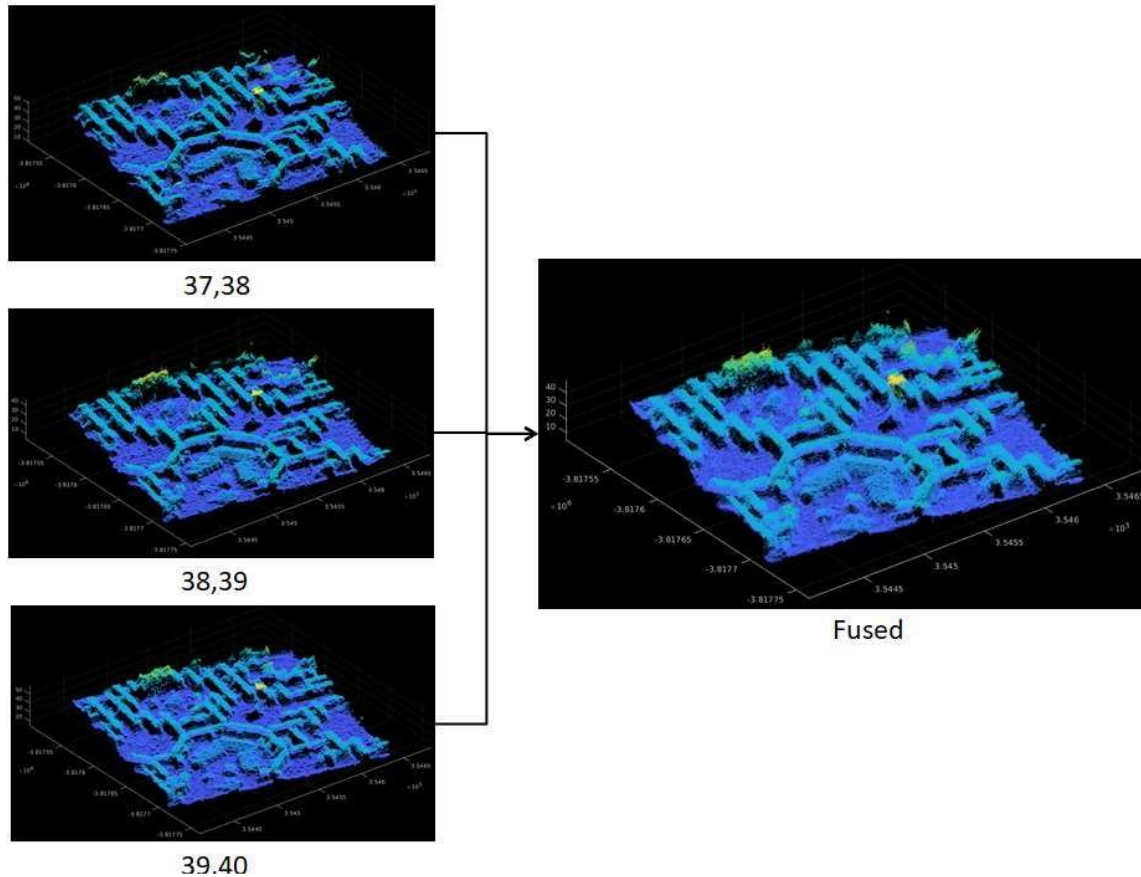
**37,38**

**38,39**

**39,40**

**Fused**

Figure 5: Fused Point Cloud Registration

*3.5 Point Registration and Fusion –*
Many point clouds can be created from any pairs of images. The point clouds can be merged into a single point cloud using Matlab. Matlab contains a Matlab Engine API for Python which is a Python package that can be used to call Matlab functions from Python. Once the Matlab engine is imported into Python, a function can be used to start the engine which creates a Matlab object and it can be attached to a new Matlab process. Matlab uses the ICP algorithm to register two point clouds. One of the two point clouds to be fused is considered to be a moving point cloud while the other is considered to be a fixed point cloud. The moving point cloud is registered according to the fixed point cloud and a rigid transformation is returned. An aligned point cloud object as well as the root mean square error can also be received as output. The root means the square error is returned as the Euclidean distance between the aligned point clouds.

ICP algorithm is an iterative process and is based on the concept of Iterative Closest Point. Each point on one point cloud is associated with the closest point on the second point cloud. The point clouds that are input can optionally be removed of outliers and down sampled to increase accuracy. The ICP algorithm consists of four steps. In the first step, the points between the moving and fixed point cloud are matched. Then the incorrect point matches are removed (outlier filter). The third step involves rotation and translation of the moving point cloud so that the distance between corresponding points is minimal. The fourth step checks if the algorithm has reached required accuracy and it stops, otherwise the four steps are repeated again.

The ICP algorithm is applied until the RMSE value of the point cloud pair is minimal. The pair of point clouds are then merged using a Matlab which runs a grid box over the overlapping regions and the points within the regions are averaged based on their locations, colors and normals. The fused point cloud can then be again aligned with more point clouds produced by better image pairs.

## IV. EXPERIMENT AND RESULT

This section provides results for the 3D reconstruction of a given set of satellite data. As mentioned earlier, the dataset contains 50 DigitalGlobe WorldView-3 panchromatic images from its sample set of publicly available

images. The images are a collection of images of the area from different angles. The task was to create a 3D image of the specified area by considering all the images. We select some pair of images whose variance is optimum (i.e. not too high or too low) and also close to each other. Point clouds of these pairs are aligned and fused together to obtain a better result.

Here, we have a collection of 50 images whose indexing is done from 0 to 49. We take 3 pair of images i.e. 37, 38; 38, 39 and 39, 40 whose variance matches the criteria mentioned above. Figure 5 shows the point clouds of three images and the point cloud obtained after point registration and merging.

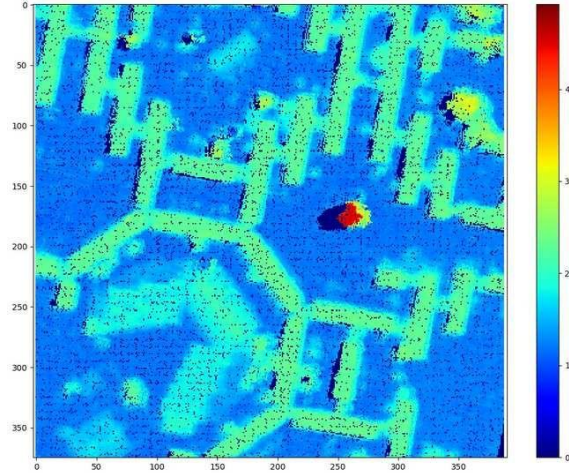DEM generated from this fused point cloud is shown in figure 6.



Figure 6: DEM after point cloud registration

## V. REFERENCES

[1] Automatic 3D Reconstruction from Multi-Date Satellite Images – Gabriele Facciolo, Carlo de Franchis, Enric Meinhardt-Llopis.
[2] 3D Reconstruction from Multimodal Satellite data – LeadingIndia AI
[3] 3D City Modelling from High Resolution Satellite Images -  S Kocaman, L.Zhang, A.Gruen, D.Poli
[4] Hirschmüller, Heiko. (2011). Semi-Global Matching: Motivation, Development and Applications. 173-184.
[5] An automatic and modular stereo pipeline for pushbroom images, Carlo de Franchis, Enric Meinhardt-Llopis, Julien Michel, Jean- Michel Morel, Gabriele Facciolo. ISPRS Annals 2014.
[6] On Stereo-Rectification of Pushbroom Images – Carlo de Franchis, Enric Meinhardt-Llopis, Julien Michel, Jean-Michel Morel, Gabriele Facciolo, ICIP 2014.