GENETIC ALGORITHM BASED STEREO IMAGE CORRESPONDENCE USING MULTI-OBJECTIVE FITNESS FUNCTION FOR REMOTELY SENSED IMAGES

Manimala Mahato[1], Shirish S. Gedam[2], Jyoti Joglekar[3] and B. Krishna Mohan[4]
[1]Doctoral Student, [2,4]Professor, Centre of Studies in Resources Engineering,
Indian Institute of Technology Bombay, Powai, Mumbai, India-400076
Email: [1]manimala.mahato@iitb.ac.in, [2]shirish@iitb.ac.in, [4]bkmohan@csre.iitb.ac.in
[3]Professor, DJSCOE, JVPD Scheme, Mumbai, India-400056
[3]Email: jyotij1968@gmail.com

ABSTRACT

A novel dense stereo image correspondence method using genetic algorithm with multi-objective fitness function is proposed and implemented in this paper. Stereo correspondence is a core research area for many 3D model generation applications. The proposed method improves the number of inliers despite of noise, occlusion, geometric and radiometric distortion present in remotely sensed images by analyzing the stereo image pair. Here, the correspondence problem is considered as an optimization problem and is solved using population based evolutionary multi-objective optimization (EMO) strategy. In the proposed method, the genetic algorithm steps such as initialization of the population, fitness function, crossover and mutation operation are customized and implemented to solve the stereo image correspondence problem. Each candidate in the population has a 2D structure representing the disparity map for the input stereo image pair. To solve this complex real-world optimization problem of stereo correspondence, improved initialization of population is used instead of randomly selected population, because it leads to more accurate and faster convergence. To initialize the populations, the SIFT feature descriptor for each pixel is computed and matching is performed using the similarity measures namely, Euclidean distance between the descriptors and Spectral Angle Mapper (SAM). The most suitable disparity maps are chosen based on the evaluated parameter values using the designed fitness functions considering the constraints related to stereo image pair such as epipolar constraint, which encodes the epipolar geometry, a similarity measure which is useful to decide accuracy of the corresponding points. Here, two objective functions used are: number of inliers computed using the fundamental matrix and an energy minimization function considering discontinuities and occlusions. To demonstrate the effectiveness of the proposed approach, the results are obtained by applying the proposed method on a remotely sensed stereo image pair.

Keywords: Dense correspondence, multi-objective fitness function, genetic algorithm, stereo matching, evolutionary optimization.

INTRODUCTION

In computer vision and computer graphics, one of the necessary tasks is the process of capturing the shape and appearance of real objects. This process is known as 3D reconstruction. Classical approaches for 3D reconstruction are image-based i.e. estimating structure from stereo image pairs or from image sequences. If multiple images of a real object is captured from different viewpoints, the projected locations of the same physical point in space, will be different on the captured images. The difference in the projected locations is used to infer the depth information. For structure from stereo, a single pair of images of the same object or scene is used. In general, two cameras located at two different spatial locations with different orientations capture the stereo image pair simultaneously. Human being can perceive depth by comparing two slightly displaced images of the observed scene. But the imitation of this capability of perceiving depth by computational techniques remains to be a challenging and unsolved problem.

The problem of 3D reconstruction from two or more images is divided into two sub-problems: feature correspondence and structure estimation. To explain the first sub-problem, considering, images A and B of the same physical 3D point X are captured from two different viewpoints. The projected location of point X on image A is point P and on image B is point P’. Point P and P’ are known as correspondence point to each other. The shifting of position of point P to P’ is known as disparity. Disparity is the prime input for depth calculation. Hence, feature correspondence problem is to find the exact location of point P’ on image B given the location of point P on image A. The issue with the structure estimation problem depends on the amount of a priori information available i.e. the intrinsic and extrinsic parameters of the camera(s). However, even if these parameters are known, the accuracy of the reconstructed structure is affected by the accuracy achieved by the correspondence problem. As a consequence, the challenge remains for developing algorithms to solve the correspondence problem. Thus, the main task of stereo image analysis is to provide the dense disparity map i.e. disparity at each pixel with better accuracy, which is the basic requirement for 3D reconstruction.

In this work, a novel dense stereo image correspondence method using genetic algorithm with multi-objective fitness function is proposed. Here, the correspondence problem is considered as an optimization problem and is solved using population based evolutionary multi-objective optimization (EMO) strategy. The block diagram of the proposed method in Figure 1 shows the genetic algorithm steps such as initialization of the population, fitness function, crossover and mutation operation. To solve the stereo image matching problem, operations of genetic algorithm are iteratively implemented till the termination condition satisfied.

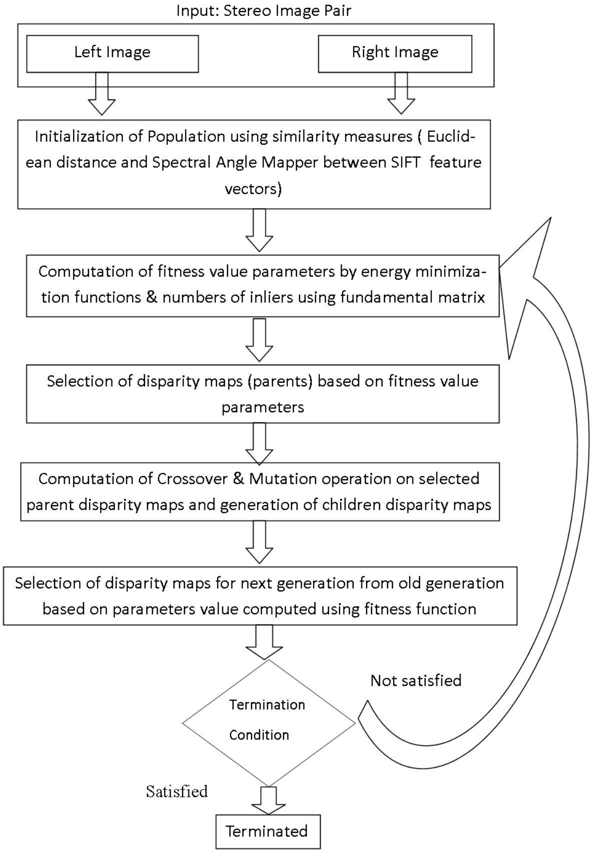


Fig. 1: Block diagram of the proposed method

In the following subsections, genetic algorithm issues related to stereo matching is discussed which includes initialization of the disparity space, the encoding mechanism for the disparity maps, the formulation of the fitness function and the appropriate crossover and mutation operators.

DENSE FEATURE DESCRIPTOR EXTRACTION

Feature detectors and descriptors

In our proposed method, the first step is to find the suitable features representing each pixel in the stereo image pair. In general, image feature points are referred as the keypoints with in the image with some specific properties such as textures, edges, corners, image gradients etc. Feature points are visually identifiable points. Feature extraction process is used for locating any specific element in many computer vision applications. While extracting features, some of the important factors to be considered are invariance, detectability, accuracy and interpretability.

Generally, feature points are sparse and represent the keypoints uniquely. These keypoints along with their unique features are used as the primary input for the further processing and analysis in many applications in the area of photogrammetry and computer vision. The invariance property of a feature ensures the detectability of the same keypoint in a stereo pair under different transformations like geometric and radiometric. Feature detectors locate the same interest point in 2D locations, in the different images uniquely. Feature descriptor is assigned to the location of that interest point after analyzing the neighborhood region of that location. Feature descriptors are used to characterize the interest point using intensity variation, change in gradient, histogram considering gradient direction and magnitude with respect to its neighboring points.

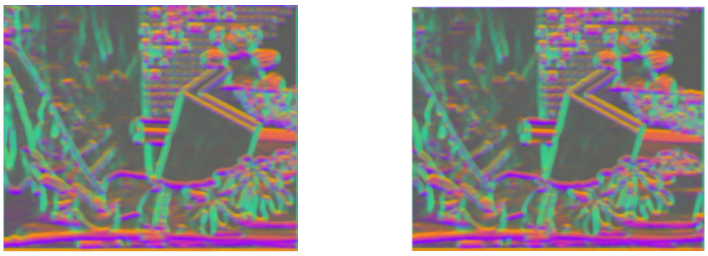
For our proposed method, we are in a need of a feature descriptor which will assign descriptors to each pixel location. Image intensity value is not the good option because it is not invariant to illumination.

Pixel wise SIFT feature descriptor

Here, pixel wise SIFT descriptor is extracted to characterize local image structures and encode contextual information by analyzing each pixel with respect to the neighboring pixel in terms of intensity variation, gradient variation, histogram of magnitude, gradient, and direction. The use of SIFT features allows robust matching invariant to scale and illumination. In [lowe, 2004], SIFT descriptor is estimated through two components: feature extraction and detection, that results sparse feature representation of the keypoints. However, in this paper, SIFT descriptor is estimated for every pixel in an image using only feature extraction component [Liu, 2011]. For every pixel in the image, the 16 x 16 neighborhood is divided into 4 x 4 cell array. Gradient magnitude and orientation is computed for each pixel using pixel difference. The orientation is quantized into 8 bins in each cell, and the SIFT representation of 4 x 4 x 8=128D vector is obtained. A Gaussian weighting function is used to assign a weight to the magnitude of each pixel. Here, the entire feature vector of 128D is used for dense matching. The effectiveness of the dense SIFT feature descriptor with respect to the image structure is demonstrated using one stereo image pair from Middlebury in Figure 2. For visualization purpose 128D SIFT feature vector is projected to 3D color space. Figure 2a and 2b is showing the left image and right image of the stereo image pair whereas Figure 2c and 2d is showing the SIFT images for the same. The visualization of the SIFT images demonstrates that, the pixels with similar local structure share the similar color. However, for computation of disparity using our proposed approach, the entire 128 dimensions are used.

Now, we have SIFT descriptors for each pixel of the input stereo image pair. Our next task is to build the disparity space by finding the similarity of these descriptors.

(a) Left Image (b) Right Image



(c) Left SIFT Image (d) Right SIFT Image

Figure 2: Visualization of SIFT images (c) and (d) of stereo image pair (a) and (b).

STEREO MATCHING USING MULTI-OBJECTIVE GENETIC ALGORITHM

In general, the computation of genetic algorithm starts with randomly generated initial population which is composed of various chromosomes. The genetic algorithm recursively generates the next generation from previous generation using crossover and mutation operator. All the chromosomes are evaluated by a fitness function. The parent chromosomes for the next generation are then selected based on the fitness value of each chromosome. Thereafter, the recursive process is continued until the stop condition is satisfied. Here, genetic algorithm is discussed in the context of stereo matching.

Disparity space generation using feature matching

After the computation of SIFT feature vector for each pixel in stereo image pair, the next step is to estimate the three-dimensional disparity space based on the similarity values of the feature vector of these pixels. For each pixel in the left image, the corresponding pixel in the right image is found using equation 1.

$$D(x,y,k) = \min\{ \sum_{\substack{-w \le i,j \le w \\ d_{min} \le d \le d_{max}}} \Psi(F_L(x+i, y+j), F_R(x+i, y-d+j))\} \tag{1}$$

Where w is the window radius, $F_L(x,y)$ and $F_R(x,y)$ are 128D feature vectors at pixel (x,y) respectively, ψ is the similarity function, k is the index number for different window sizes. Here, as a similarity measure, we use Spectral Angle Mapper (SAM) which identifies similarity with respect to the angle and Euclidean Distance (ED) which identifies similarity with respect to the magnitude of the feature vector. SAM is a very widely used technique in the field of hyperspectral image classification. The lower the value of SAM and ED signifies the higher the similarity between the feature vectors under consideration. The disparity space is filled by the disparity value for which the matching likelihood is higher. The matching likelihood of x with different pixels in the right image is ranked according to the value of angle i.e. θ. Smaller value of θ defines the higher matching likelihood. The angle and euclidean distance between the template feature vector t in the right image and the reference feature vector r in the left image is given in equation (2) and (3) respectively.

$$\theta = cos^{-1} \frac{\sum_{i=1}^{n} t_i r_i}{\sqrt{\sum_{i=1}^{n} t_i^2} \sqrt{\sum_{i=1}^{n} r_i^2}} \tag{2}$$

$$ed(t,r) = \sqrt{\sum_{i=1}^{n} (t_i - r_i)^2} \tag{3}$$

To generate the disparity space, we have used eight different window sizes 3 x 3, 5 x 5, 7 x 7, 9 x 9, 11 x 11, 13 x 13, 15 x 15, 17 x 17. Use of different window sizes in the matching process solves the problem of fixed window. Finally, eight disparity maps are estimated using each similarity measure SAM and Euclidean distance respectively. So, there will be in total sixteen different disparity maps in the disparity space. Use of two different similarity measure ensures enough diversity in the disparity space.

Encoding scheme for Disparity Maps

Encoding scheme for representing the solutions of the optimization problem is application dependent and fundamental to all genetic algorithms. In order to apply genetic algorithm in stereo matching, the type of chromosome structure should be defined as per two-dimensional image signals. Thus, 2D disparity map is simply encoded as a chromosome and disparity value for each pixel of the disparity map becomes a gene value of the chromosome.

Multi-objective Fitness Function

In order to identify the fittest individual during the evolutionary process, a function needs to assign a degree of fitness to each chromosome in every generation. This function, known as the fitness or evaluation function, controls the evaluation process. Here, the fitness of each disparity map is evaluated based on a multi-objective function. In the literature of stereo image analysis, in general, the optimal disparity map is evaluated by the compatibility between the corresponding points and continuity of the disparity map. This criteria can be employed by an energy function. However, the energy criteria does not accept outliers like occluded pixels and discontinuity at edge pixels because the energy is increased due to these outliers. Hence, we are allowing these outliers by adding one more evaluation criteria with the energy function. The second criteria considers the constraints related to stereo image pair such as epipolar constraint, which encodes the epipolar geometry. This criteria basically counts the number of inliers using the fundamental matrix. We employ these two stereo matching criteria, which is the energy minimization criterion and maximizing the number of inliers, as objective functions to compute the fitness of each chromosome in the evolutionary optimization algorithm.

Each chromosome is evaluated by determining the energy representing the compatibility between corresponding pixels in the stereo image pair and the continuity in the generated disparity map. The equation of the energy function is as follows:

$$E = \sum_{x,y} \left|(F_L(x,y) - (F_R(x + d_{x,y}, y))\right| + \sum_{x,y} (\left|(d_{x,y} - d_{x+1,y})\right| + \left|(d_{x,y} - d_{x,y+1})\right| \qquad (4)$$

Where $F_L(x,y)$ and $F_R(x,y)$ are SIFT vector of left image and right image respectively in stereo image pair, $d_{(x,y)}$ is the disparity map, and λ is a weighting factor. The first term and second term represent the energy values of compatibility and continuity, respectively. The disparity map with smaller energy value having larger fitness value.

The second fitness value is defined by maximizing the number of matching points, which is known as inliers. For the disparity map under consideration, number of inliers are computed using the fundamental matrix, respecting the epipolar constraint of stereo matching. The fundamental matrix is computed using RanSAC algorithm [Fischler et al., 1981].

Generation of Initial Population

To start with, genetic algorithm needs the initial population consists of various initial disparity maps. Here, two very important factors which control the convergence of the algorithm is the population size and how to generate the initial population. Population size specifies the number of individuals in each generation. With a large population size, the genetic algorithm searches the solution space more thoroughly, thereby reducing the chance that the algorithm returns a local minimum. However, a large population size also causes the longer computation time. On the other hand, too small population size may result premature convergence without finding an appropriate solution. In this paper, to avoid the problem, empirically the population size is set to be 90.

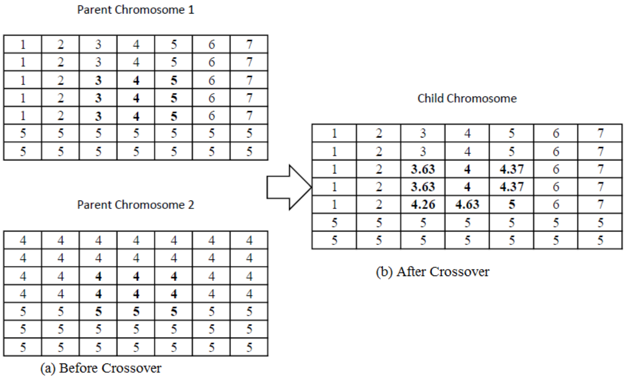
Most genetic algorithms use a random function to generate the initial population, accordingly, the convergence speed is too slow. To improve the convergence, it is better to start with an informed population, in which a higher selected possibility of disparity is assigned to the chromosomes. Using the disparity space, the solutions are seeded in the areas where optimal solutions are likely to be found. The initial disparity maps are generated by selecting the disparity value for each pixel from the disparity space on random basis. The method of generating the initial population is known to affect the convergence of the problem.

Crossover Operator

Genetic operators such as crossover and mutation provide the basic search mechanism in GA and serve as a method to continually improve the fitness and converge to an optimal solution. In our scenario, the crossover operation creates one child chromosome by taking a weighted average of two parent chromosomes using equation (5). The weight is specified by a randomly generated parameter, *Scale,* within the range [0,1].

$$child = parent1 + scale * (parent2 - parent1) \tag{5}$$

An example crossover operation is shown in Figure 3 with parent 1 and parent2 of size 7 x 7. Figure (3a) shows the parent chromosomes before the crossover operation and Figure (3b) shows the child chromosome computed after crossover operation. After crossover, the newly generated disparity map may have a lower energy and better number of inliers than either of its parents.



Parent Chromosome 1

1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	**3**	**4**	**5**	6	7
1	2	**3**	**4**	**5**	6	7
1	2	**3**	**4**	**5**	6	7
5	5	5	5	5	5	5
5	5	5	5	5	5	5

Parent Chromosome 2

4	4	4	4	4	4	4
4	4	4	4	4	4	4
4	4	**4**	**4**	**4**	4	4
4	4	**4**	**4**	**4**	4	4
5	5	**5**	**5**	**5**	5	5
5	5	5	5	5	5	5
5	5	5	5	5	5	5

(a) Before Crossover

Child Chromosome

1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	**3.63**	**4**	**4.37**	6	7
1	2	**3.63**	**4**	**4.37**	6	7
1	2	**4.26**	**4.63**	**5**	6	7
5	5	5	5	5	5	5
5	5	5	5	5	5	5

(b) After Crossover

Fig 3: Crossover operation in the proposed method

Mutation Operator

Mutation operator is the process of changing the genes randomly by supplying new genetic material. The crossover operator generates one or more new child of the modified population by considering the location of the parent chromosomes. Mutation operator, a unary operator, allows an EO to search locally around a solution independent of the location of other solutions in the population. In our proposed method for stereo matching, disparity smoothness constraint is considered through mutation operator. This is done by considering the disparity value of the pixel under consideration and the two neighboring disparity values using equation (6).

$$leftDiff = d(x, y-1) - d(x, y);$$

$$rightDiff = d(x, y+1) - d(x, y);$$

$$d(x, y) = -0.5 + \frac{leftDiff}{leftDiff + rightDiff}; \quad if\ leftDiff \leq rightDiff \tag{6}$$

$$d(x, y) = -0.5 + \frac{rightDiff}{leftDiff + rightDiff}; \quad else$$

Minimization Process

The genetic algorithm is implemented as an iterative procedure after addressing the above issues. The population is evolved from one generation to next generation through each iteration. In each generation, crossover and mutation operations are processed on two randomly picked chromosomes. While selecting chromosomes for the next generation, the *elitist strategy* [Goldberg, 1989] is applied. This strategy ensures the fitness of the worst chromosome, in each generation, will not decrease. The above process is repeated until the termination condition is satisfied. In our approach, the iteration will be terminated if the relative change of the best chromosome over the generations is less than the tolerance threshold.

RESULTS AND ANALYSIS

To demonstrate the effectiveness of the proposed approach, the results are obtained by applying the proposed method on a remotely sensed stereo image pair. The stereo image pair is a subsection of the remote sensed image of Nahan village on top of hill near Simla, India. The stereo image pair is taken by IRS 1C having spatial resolution of

5.8 m in Panchromatic. Its size is 254 X 254 pixels. Substantial illumination difference and geometric distortion among the image pair is present. The left image and right image of the dataset is shown in Figure 4.

After computing the SIFT image of the input stereo image pair, the disparity space is generated using two similarity measures namely spectral angle mapper and euclidean distance with eight different window sizes. The disparity spaces are shown in Figure 5. There are eight different disparity maps within each disparity space. The size of each disparity space is 254 X 254 X 8 for the given input stereo image pair. Though the disparity maps, in the disparity spaces, are looking very similar, they are different with each other at the non-textured areas where prominent features are not there and the distorted area due to noise, occlusion, geometric and radiometric distortion.

Two example chromosomes of initial population, generated from two disparity maps, are shown in Figure 6. Starting with the initial population, genetic algorithm iteratively evolves the disparity maps by each generation. Here, the evolutionary process is guided by the two objective functions which checks the energy and number of inliers of each disparity map. After termination of the genetic algorithm, the disparity map with minimum energy and maximum inliers among the population, is termed as the final output of our proposed algorithm. Figure 7 is showing the 3D surface plot of the disparity map, computed using our proposed method. The proportion of inliers in the computed disparity map using our proposed method is 76.68% whereas in the initial population the range for the same was 66% to 71%.

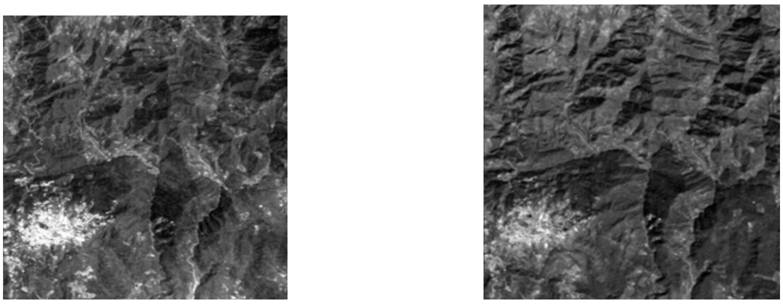


Figure 4: Remotely sensed stereo image pair (Nahan village): Left Image and Right Image

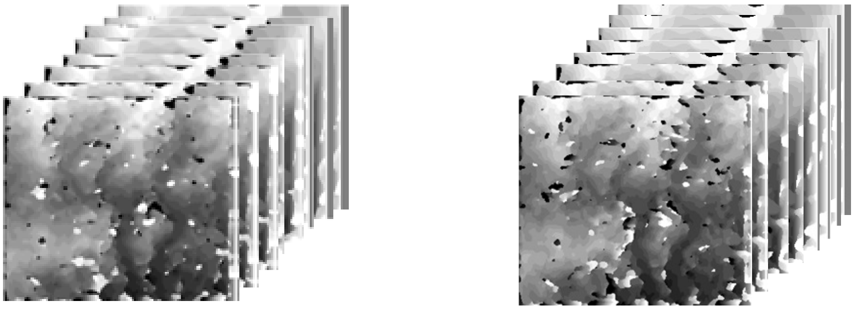


Figure 5: Disparity space using (a) Spectral Angle Mapper (SAM) (b) Euclidean Distance (ED)

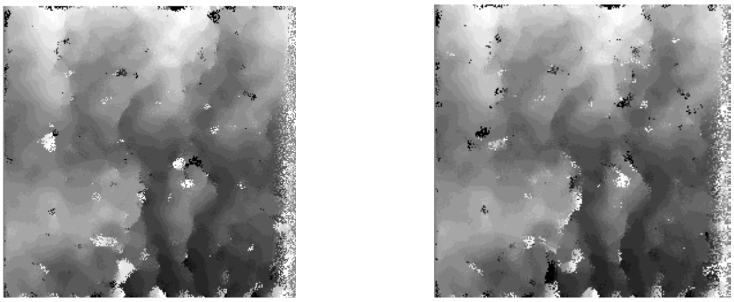


Figure 6: Example chromosomes from Disparity spaces using (a) SAM and (b) ED

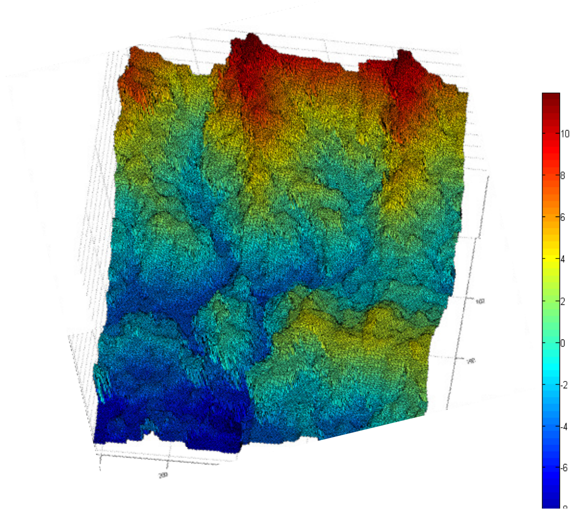


Figure 7: 3D surface plot for satellite image of Nahan village (near Simla, India) by IRC 1C

CONCLUSION

The dense disparity map obtained by our proposed novel stereo image correspondence method using genetic algorithm with multi-objective fitness function is useful for accurate 3D reconstruction. 3D reconstruction using remotely sensed stereo image pair remains a challenging task due to noise, occlusion, geometric and radiometric distortion. Our proposed method improves the number of inliers despite of the above-mentioned problems. The stereo correspondence problem is solved by genetic algorithm using multi-objective fitness function by considering the problem as optimization problem. Various operations of genetic algorithm are designed in such a way that it can solve the complex real-world optimization problem of stereo correspondence. Our genetic algorithm achieves faster convergence due to the improved initialization of population. Various constraints related to stereo image pair is encoded in the multi-objective fitness functions which guided the genetic algorithm towards the optimum disparity map. The effectiveness of the proposed method is demonstrated using a remotely sensed stereo image pair with various challenging characteristics.

REFERENCES

Blake A. and Zisserman A.1987, *Visual reconstruction*, volume 2. MIT press Cambridge.

Faugeras O. 1993, *Three-dimensional computer vision: a geometric viewpoint*. MIT press.

Fischler, M. A. and Bolles, R. C. 1981, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* , Vol. 24, ACM, pp. 381-395.

Goldberg, D. E. 1989, *Genetic algorithms in search, optimization, and machine learning*, Reading: Addison-Wesley

Konak, A., Coit, D. W. and Smith, A. E., 2006, Multi-objective optimization using genetic algorithms: A tutorial, *Reliability Engineering & System Safety*, Vol. 91, Elsevier, pp. 992-1007.

Liu, C., Yuen, J. and Torralba, A., 2011. Sift flow: Dense correspondence across scenes and its applications. *IEEE transactions on pattern analysis and machine intelligence*, *33*(5), pp.978-994.

Lowe, D., 2004, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision,* Vol.60(2), pp. 91-110.

Schaffer, J. D., 1985, Multiple objective optimization with vector evaluated genetic algorithm, in *Proceeding of the First International Conference of Genetic Algorithms and Their Application*,pp.93-100.

Senthilnath, J., Kalro, N. P. and Benediktsson, J. A., 2014, Accurate point matching based on multi-objective genetic algorithm for multi-sensor satellite imagery, Applied Mathematics and Computation, Vol. 236, Elsevier, pp. 546-564.

Srinivas, N. and Deb, K., 1994, Multi-objective optimization using non-dominated sorting in genetic algorithms, *Evolutionary computation*, Vol. 2, MIT Press, pp. 221-248.