Volume-04
Issue-03
March -2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Analysis, Implementation and Application of Stereo Matching Algorithm for Image Depth Understanding

[1]Siddiqui Tamanna Ashraf & [2]Kakar Varun Kumar

[1]M. Tech. Scholar, Department of E & C Engineering, B. T. Kumaon Institute of Technology, Dwarahat, Uttarakhand (India)
[2]Assistant Professor, Department of E & C Engineering, B. T. Kumaon Institute of Technology, Dwarahat, Uttarakhand (India)

ARTICLE DETAILS

Article History
Published Online: 13 March 2019

Keywords
Stereo matching, Stereo matching algorithm, depth map, SAD, HSAD, SURF, Hessian Matrix, HSAD-Hessian-OpenSURF.

ABSTRACT

Stereo vision, resulting in the knowledge of depth information in a scene, is of significant importance in the field of machine vision, robotics and image analysis. Stereo matching is the process of computing a disparity map from a pair of stereo images. At the heart of every stereo matching algorithm is a solution to the correspondence problem – the problem of finding points in the right and left image that correspond to a single point in the real world. The paper presents a new method of feature based stereo matching algorithm. The state of the art algorithm for hybrid segmentation (HSAD) works fast and produce highly accurate depth map, but it is sensitive to different image conditions (i.e. illumination, rotation, blurring, scale, clutter, etc.). To overcome the problem, this research proposes an algorithm that utilizes OpenSURF algorithm and Hessian Matrix Laplace Function to enhance the image quality of HSAD matching. Speeded-Up Robust Feature (SURF) is a fast and performing scale and rotation-invariant interest point detector and descriptor. It uses a Hessian matrix for blob interest point extraction. Hessian matrix minimizes the filtering process and speed up the search for correspondence. The hybrid technique is integrated with the SAD stereo matching algorithm to determine the disparity estimate of each image pixel.

1. Introduction

The light rays projected onto the retina present our visual system with an image of the world that is inherently two-dimensional, yet we can interact with the three-dimensional world, even in new situations, or with unknown objects. That we accomplish this task easily implies that one of the functions of the human visual system is to reconstruct a 3-D representation of the world from its 2-D projection onto our eyes. The very idea of stereovision directly comes from the human eyes.

A stereo setup built from two cameras can give information about depth. Depth is in fact the variable which gives the 3D Perception as shown in Figure 1.1. Our eyes receive horizontal deviated views, and our brain automatically analyzes the disparities of objects and fuses them to 3D perception.

There are several ways to acquire depth information for a 2D image. One way is to use a structured light sensor or time-of-flight. However, these active sensors suffer from various drawbacks. They are subjected to systematic errors such as noise and ambiguity, due to a particular sensor used.

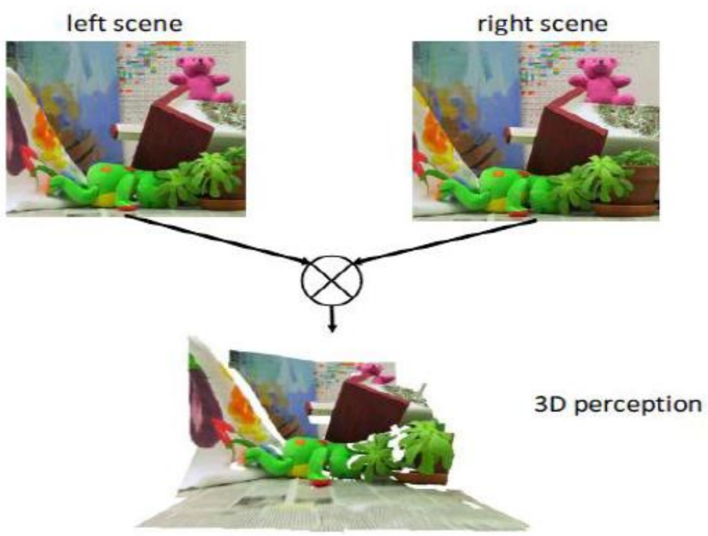

Figure: 1 3D Perception

They also suffer from non-systematic errors such as scattering and motion blur. Moreover, these active sensors perform satisfactorily only up-to 5-7 meters of distance, which makes them too sensitive to be used in outdoor environments. Another way is to use a stereo image pair in combination with triangulation. The latter technique is normally referred to as stereo vision, stereo matching or stereo correspondence. Stereo vision sensors are passive sensors. They are more reliable and robust. They produce disparity maps of high resolution. They are suitable for both indoor and outdoor environments. In principle, stereo matching is about finding matching pixels in two given input images taken from slightly horizontally separated points

2. Related Work

Tombari et al. (2008) [2] proposed a cost aggregation strategy based on segmentation to improve the results of fast stereo matching algorithms. The proposed aggregation strategy is deployed by a global method using Belief Propagation (BP) or local method using sum of truncated absolute difference (TAD) correlation measure for RGB colour images. Mean shift is used for performing segmentation which accurately segments the images but is not very fast and thus, attains near-real time performance. The algorithm can process 5fps for 320× 240 pixels and 16 disparity levels for standard Tsukuba pair. The result shows that the proposed algorithm is two times faster than the variable window and more than three times faster than segmentation based methods.

Alagoz (2008) [3] proposed a region based stereo matching algorithm to obtain depth map from two colour images of the same scene, taken from two stereo cameras. Two algorithms have been developed on the basis of region stereo matching algorithm. The first method is Global Error Energy Minimization by Smoothing Functions. The average filter is then applied repeatedly on the calculated energy matrix to obtain smooth disparity. The disparity with minimum error energy is chosen as most reliable disparity value for a particular pixel of disparity map. The second method is Line Growing Based stereo matching. The method first selects the root point and then grow region for the selected root point. The results obtained by first method are more reliable and smoother, but is slower to be implemented on software. The second method is faster for software implementations. Application of large sized median filtering makes the disparity and depth map smoother in intelligent systems such as robot navigation.

Juan et al. (2009) [4] compares the three robust feature detection methods: Scale Invariat Feature Transform (SIFT), Principal Component Analysis (PCA)–SIFT and Speeded Up Robust Features (SURF). The applicability of these methods in recognition is analyzed by using KNN (K-Nearest Neighbour) and Random Sample Consensus (RANSAC). KNN is used for finding the matches, and RANSAC for rejecting inconsistent matches from which the inliers can take as correct matches.

The three methods performance is compared for scale changes, rotation, blur illumination changes and affine transformations. The experimental results shows that SIFT and SURF methods provide good performance, but SURF is faster. PCA-SIFT show its benefits in rotation and illumination changes.

Tombari et al. (2010) [5] proposed a memory- efficient dense stereo matching algorithm. The stereo matching algorithms (including both global and local methods) are first selected on the basis of speed and low memory requirement, and then compared in terms of accuracy and speed. The comparison is done on both standard benchmarking dataset and novel dataset. The result shows that the global methods such as scan line optimization (SO) and dynamic programming (DP) provides good accuracy but is less efficient than local methods such as 3windows, variable windows and multiple windows. The result depicts that the novel dataset images are affected by the artificial noise. It is also noted that the accuracy of stereo matching algorithms is slightly improved at a notable increase in computational cost. The DP is 1.2 times more accurate and 2.2 times slower than the (Sum-of-Absolute Difference) SAD algorithm. Comparatively, AWSO (Adaptive Window Scanline Optimization) is 1.3 times more accurate than the SAD algorithm, but 22.6 times slower.

A correlation based dense stereo matching algorithm is presented by **Fradi et al. (2011)** [6] for determining matching pixels in challenging areas such as occluded regions, textureless areas and along discontinuities. The occluded values are filled by including edge detection, preventing window from containing more than one object. A variable window size with adaptive shape is selected for getting accurate results at depth discontinuities and in homogeneous areas and at the same time, keeping the complexity of the entire system low.

Sun et al. (2011) [7] proposed a propagation based stereo matching algorithm for producing dense disparity maps. From the initial disparity maps, errors are removed with seed pixel detection, scan line propagation and proper refinement. The proposed algorithm using pixel wise line segments work well along textureless regions and depth discontinuities. The algorithm does not require segmentation operations and computations are performed in each pixel’s line segment resulting in low computational complexity. The line segments are drawn with colour and connectivity constraint. The result shows that the proposed algorithm gives good performance both in terms of speed and accuracy. The processing times for the four stereo pairs (Tsukuba, Venus, Teddy, and Cones) are 2.5 seconds, 3.8 seconds, 8.7 seconds and 8.6 seconds respectively.

Nalpantidis et al. (2011) [8] proposed efficient hierarchical based stereo matching algorithm that can satisfy the demands of self-directed outdoor robotics applications. The algorithm is presented to estimate disparity, taking ideas from motion estimation method that was first developed to code videos only. The DXD search blocks are used to find matching pixels between two stereo images, thus reducing the space for correspondence search. The result of this core is refined by Gaussian weighted aggregation and 3-D cellular automata. Further, parallelization and pipelining of the chosen modules in the proposed approach result in hardware implementation, making it suitable for real outdoor robotics applications. The refining methods used make the proposed algorithm complex, but the results obtained are of high quality. The performance of the proposed algorithm is not affected by lens distortion, imperfect lightning conditions and spatial displacements in input stereo images. The proposed algorithm can evaluate uncalibrated and non-rectified image pairs quickly and with good quality, thus making it suitable for real-time outdoor robotics applications. Both standard and novel image sets have been used for evaluating the performance of the proposed algorithm.

Kamencay et al. (2012) [9] proposed a segmentation based stereo matching algorithm. The two algorithms, belief propagation and mean shift segmentation are combined to make a hybrid algorithm. The hybrid algorithm takes the benefit of both algorithms. Belief propagation segments the images very accurately and mean shift algorithm give fast results. SSD (Sum-of-Squared Difference) and SAD (Sum-of-Absolute Difference) algorithms are compared for selection on the basis of depth map estimation. SAD is selected because it is much faster than SSD. The result shows that the depth maps obtained by SAD algorithm using hybrid segmentation are much better than those obtained by SAD algorithm without segmentation. The run time for SAD algorithm is 105 seconds and for proposed Hybrid method is 29 sec. The proposed Hybrid Sum-of-Absolute Difference (HSAD) algorithm provides good accuracy with near-real-time performance.

Yang (2012) [10] examined the existing local aggregation method and proposed a non-local aggregation method for stereo matching. The study revealed that the existing local aggregation methods are critically influenced by the local characteristics of existing window-based cost aggregation methods and are exposed to lack of texture. In this paper, the calculated values of matching cost are aggregated intelligently on the basis of similarity of pixels on a Minimum Spanning Tree (MST) obtained from the stereo images for preserving depth edges. The image pixels are considered as tree nodes and all the edges between pixels of closest neighbours are the edges. The smallest distance between two pixels on the tree determines whether the two pixels are similar on not. The presented method is non-local because all the nodes in the tree get support from each other. The performance of proposed non- local method is best above all the local methods on the standard benchmarking site.

Jiao et al. (2014) [11] proposes a cost-volume filtering based technique that aims at removing some notable outliers still present in the final disparity maps of local stereo matching algorithms.

Two strategies have been proposed to improve the results of local stereo matching. Firstly, the performance of matching cost is improved by combining truncated absolute difference of colour, gradients and a modified colour census transform. Secondly, the secondary refinement approach is applied after the traditional disparity refinement for further refinement of results. The proposed method uses symmetric guided filter for cost aggregation. The secondary disparity refinement method is called Remaining Artefacts Detection and Refinement (RADAR). Experimental results are obtained for real world sequences and some depth-based applications for showing the potential of the method. The processing time for Tsukuba, Venus, Teddy, and Cones are 3.11s, 5.70s, 12.34s, and 12.48s respectively.

Yang et al. (2015) [12] proposed an improved Adaptive Support Window (ASW) for stereo matching algorithm. The ASW is the most popular window based approach to match pixels. The disparity of a pixel is calculated independently within a local support window. ASW determines support-weights of an anchor pixel on the basis of colour similarity and geometric proximity. However, ASW does not obtain satisfactory results in areas having repetitive pattern, occlusions and similar colour. The proposed method provides improved ASW by including many local image features. Firstly, an adaptive cross-based support window is used instead of the fixed square window to increase the accuracy of matching pixels in case of a sparse-textured area. Secondly, canny edges are used in dense-textured regions for matching pixels in corresponding images. Thirdly, a smoothness constraint on disparity continuity is imposed to reduce the errors on object boundaries. Fourthly, a disparity-refinement process is performed on the independent pixels using LRC. The result shows that the proposed method produces stable and efficient results with only a slight increase in computation time. The PC platform is 3.4 GHz Intel i7-3770 CPU. The result shows that proposed method improves the accuracy of ASW with only 10% to 36% more computation time.

3. Proposed Work

The objective of the proposed work is to develop a technique which is invariant to geometrical or photometric transformations in images and is able to establish correspondences between a pair of images taken from different viewpoints.

The proposed technique for stereo matching starts with acquisition of images, which are calibrated and rectified. This algorithm consists of the following stages:

i. Image acquisition,
ii. Epipolar geometry and image rectification,
iii. Feature Detection,
iv. Stereo matching algorithm,
v. Depth map estimation

First, radial and tangential lens distortion are removed by camera calibration, which gets intrinsic and extrinsic camera parameters. Knowledge of the camera parameters is utilized to rectify both images. After rectification, the image depth points are detected by using the proposed hybrid technique. Finally, stereo matching algorithm is applied on the left and right images with the aim to find all correspondences (matching points) and assign disparity to each image pixel. Output of the stereo matching algorithm is the disparity map.

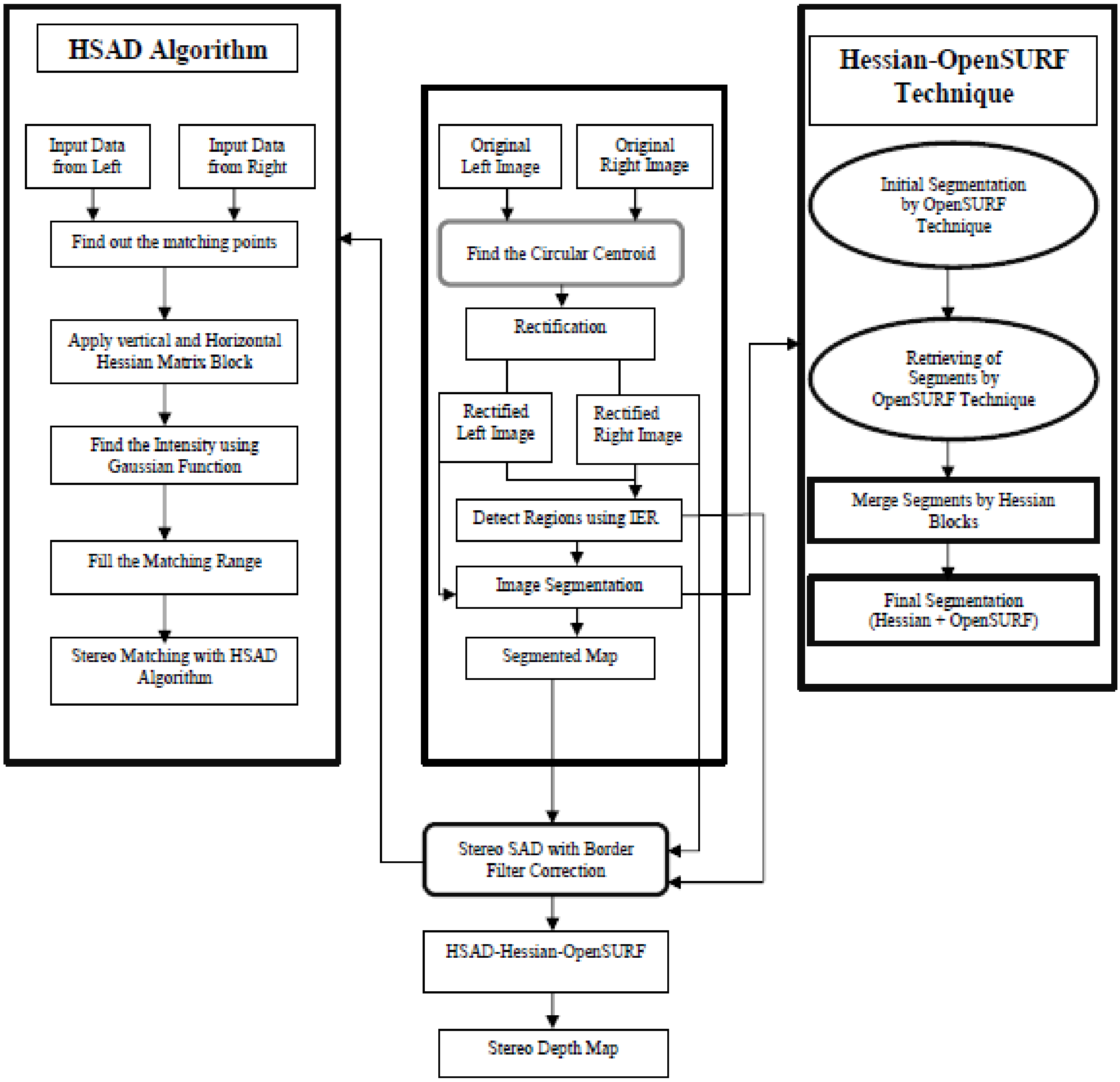

Figure: 2 Proposed approach using HSAD-Hessian Open SURF Technique

4. Results & Discussions

The proposed hybrid technique (HSAD-Hessian-OpenSURF) is rotation invariant unlike the previously proposed hybrid segmentation algorithm (HSAD) [15]. The HSAD algorithm has therefore the disadvantage of not designing the different angular depth map and different position sets of depth map that will generate recall. The recall generated, reduces the reliability factor of the depth map which is the drawback.

With the increase in the number of developed algorithms for stereo matching, evaluation criterion for studying of stereo matching is required. The criterion used for comparing stereo matching algorithms presented here, is based on computing precision, recall, F-measure and accuracy. These three parameters determine the algorithms efficiency. The definition of precision (*P*), recall (*R*) and F-measure (*F-1*) is given by:

$$P = \frac{C}{C+F} \quad (1)$$

$$R = \frac{C}{C+M} \quad (2)$$

$$F-1 = \frac{2PR}{P+R} \quad (3)$$

Where *C* is the number of correct depth points, *F* is the number of false depth points and M is the number of undetected depth points.

The precision (*P*) is measured as the number of correct depth points out of all the depth points (correct and false). The Recall (*R*) is measured as the number of correct depth points out of all the correct depth points taken into consideration. F-measure is a standard way to balance precision and recall, which is the geometric mean of recall and precision. It is in high values if both precision and recall have high values and on the other hand, if one of them has low value, the value of the F-measure is going down. A balance between these two needs to be achieved and to achieve this and to compare performance, the precision-recall curve comes in handy.

The precision-recall curves for two algorithms are shown in Figure 3. Depending on the requirement (high precision at the cost of recall, or high recall with lower precision), an appropriate algorithm can be chosen.

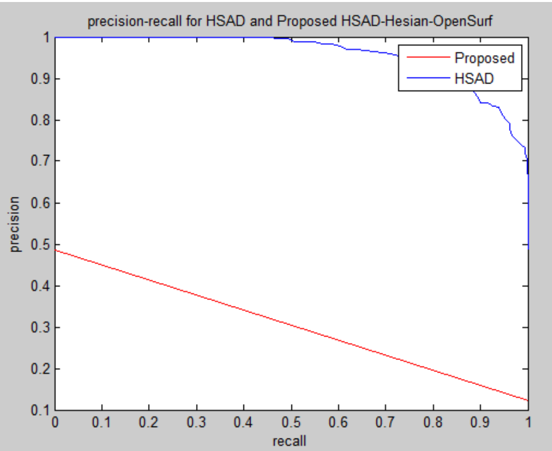

Figure: 3 Precision-recall curves for HSAD and proposed HSAD-Hessian- OpenSURF.

F-measure comes from Information Retrieval (IR) where Recall is the frequency with which correct Depth point is retrieved by an algorithm. Precision is the frequency with which retrieved Depth points are correct, and is properly a form of accuracy. F-measure is intended to combine these into a single measure of search called effectiveness.

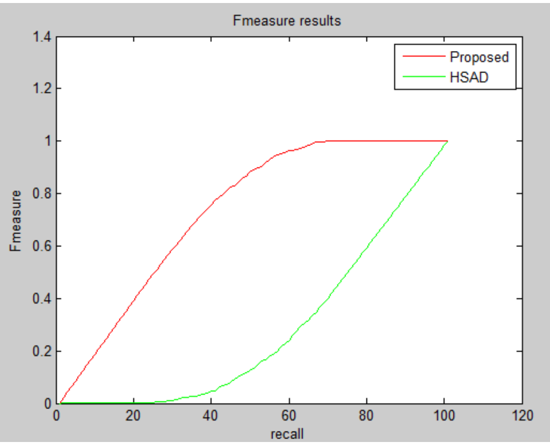

Figure: 4 F-measure results for HSAD Algorithm and Proposed HSAD Hessian – OpenSURF Technique

The figure 5 shows the k value which denotes the number of strongest Depth points. The graph shows how much accurate a particular algorithm is for finding the same number of correct Depth points. It is noted from Figure 5 that the proposed HSAD-Hessian-OpenSURF technique is 97% accurate, whereas the HSAD algorithm is only 30 % accurate.

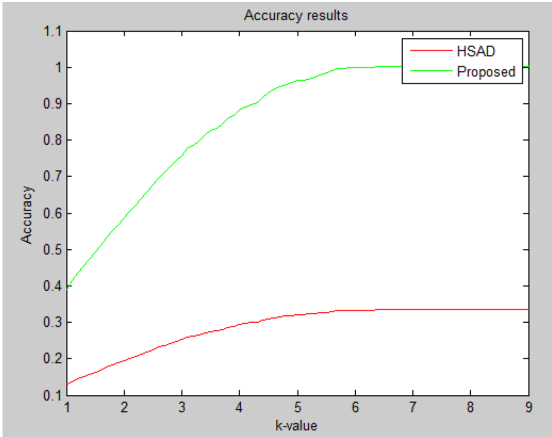

Figure: 5 Accuracy results for HSAD Algorithm and Proposed HSADHessian -OPEN SURF Technique

The comparison among the two algorithms and the simulation results show that the proposed HSAD- Hessian-OpenSURF technique is far efficient than hybrid HSAD algorithm.

5. Conclusion

The stereo matching algorithm was designed and implemented as required. We have introduced an efficient and robust HSAD-Hessian- OpenSURF technique which provides depth information under different image conditions (i.e. illumination, rotation, blurring, scale, clutter, etc.). HSAD-Hessian- OpenSURF technique outperforms the hybrid segmentation algorithm, both in speed and accuracy. We have reached 97% accuracy in depth estimation in average. The average processing time for a stereo image pair (*ART*) is reported 9 sec. Thus it is more suitable technique for real-time application system. In future work, we plan to perform experiments (we could speed up computation time and improve precision of hybrid technique) and also tests of more complex algorithms on greater number of real images with aim to compare the presented approach with other existing algorithms.

References

1. Zitnick, C., & Kanade, T. (1999). A Cooperative Algorithm for Stereo Matching and Occulsion. In tech. refort CMU-RI-TR-99-35. Robotics Institute, Carnegie Mellon University
2. Tombari, F., Mattoccia, S., Stefano, L., & Addimanda, E. (2008). Near real-time stereo based on effective cost aggregation, nternational conference on Pattern Recognition (pp. 1– 4, 10, 24)
3. Alagoz, B.B. (2008). Obtaining Depth Maps from Colour Images by Region Based Stereo Matching Algorithms. OncuBilim Algorithm and Systems Labs (Vol. 08, Art.No.04)
4. Juan, L., & Gwun, O. (2009). A Comparison of SIFT, PCA-SIFT and SURF, International Journal of Image Processing (Vol.3, Issue (4)).
5. Tombari, F., Mattoccia, S., & Stefano, L. (2010). Stereo for Robots: Quantitative Evaluation of Efficient and Low-memory Dense Stereo Algorithms, IEEE Int. Conference on Control, Automation, Robotics and Vision. Singapore.
6. Fradi, H., & Dugelay, J. (2011), improved depth map estimation in Stereo Vision. Article, EURECOM. Sophia Antipolis, France

7. Sun, X., Mei, X., Jiao, S., Zhou, M., & Wang, H. (2011). Stereo Matching with Reliable Disparity Propagation, IEEE International Conference on 3D Imaging, Modelling, Processing, Visualization, and Transmission
8. Nalpantidis, L., Amanatiadis, A., & Sirakoulis, G. C. (2011). Efficient hierarchical matching algorithm for processing uncalibrated stereo vision images and its hardware architecture, IET Image Processing (vol. 5, no. 5, pp. 481–492)
9. Kamencay, P., Breznan, M., Jarina, R., Lukac, P., & Zachariasova, M. (2012). Improved Depth Map Estimation from Stereo Images Based on Hybrid Method, Radio Engineering (Vol. 21, No. 1)
10. Yang, Q. (2012). A non-local cost aggregation method for stereo matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1402– 1409), Providence, RI, USA
11. Jiao, J., Wang, R., Wang, W., Dong, S., Wang, Z., & Gao, W. (2014). Stereo Matching With Improved Matching Cost and Disparity Refinement
12. Yang, S., & Wang, B. (2015). Improved Adaptive-Support Stereo Matching Using Local Features, Journal of Information Science and Engineering (31, pp.1557-1575)
13. Scharstein. D., & Szeliski, R. (2003), High-Accuracy Stereo Depth Maps Using Structured Light. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (vol.1, pp. 195-202)
14. Brown, M. Z., Burschka, D., & Hager, G. D, (2008) Advances in computational stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence. (vol. 25, no. 8, pp. 993-1008)
15. Scharstein, D., Szeliski R., and Zabih R. (2001), A taxonomy and evaluation of dense two frame stereo correspondence algorithms. IEEE Workshop, Stereo and Multi-Baseline Vision