Stereo Matching using HSAD-Hessian OpenSURF Technique

Tamanna Ashraf Siddiqui[1], Akhilesh Mishra[2]
[1]*Professor & HoD, Dept. of Electronics & Communication, Suyash Inst. of Information Tech., Gorakhpur, India*
[2]*Asst. Professor, Dept. of Electronics & Communication, Suyash Inst. of Information Tech., Gorakhpur, India*

***Abstract*: Stereo matching algorithms are based on image segments estimate disparities using local window-based stereo matching methods, and subsequently refine the disparity estimates by introducing smoothness constraints and performing image segmentation. However, such algorithms are sensitive to different image conditions i.e. rotation, scale, occlusion, blurring, illumination, clutter, etc. To overcome the problem, this paper proposes an approach that incorporates the open SURF (Speeded –Up Robust Feature) algorithm to determine the disparity obtained by matching depth points between stereo image pairs. The open SURF algorithm uses a Hessian matrix for depth point detection. Hessian Matrix minimizes the filtering process and speeds up the search for correspondences in stereo image pairs. Then, the open SURF algorithm is integrated with the SAD (Sum of Absolute Differences), local block-matching stereo matching Algorithm to obtain the disparity estimate of each pixel. This HSAD-Hessian-Open SURF algorithm generates highly accurate disparity map with different image conditions. This technique gives a real time performance. Results show that the proposed algorithm is more robust against different image conditions than existing segment based algorithms. All experiments were implemented in MATLAB.**

***Keywords*: Stereo Matching, Disparity, and Depth map, Hybrid Segmentation Algorithm, SURF, SAD, Hessian Matrix, and HSAD-Hessian-Open SURF.**

1. Introduction

In the last few years there has been a growing interest in creating a realistic reconstruction of a virtual 3D scene. The ability to infer information on the 3D structure and distance of a scene from two or more images taken from different viewpoints is called stereo vision [1]. The very idea of stereo vision directly comes from the human eyes. Our eyes take horizontal displaced views, and our brain analyzes the differences or disparities of objects in an inherent way and fuses them to 3D vision.

Stereo vision is generally realized through stereo matching in computer vision. Two cameras capture images of the same 3D world scene, but they are separated by a distance-exactly like our eyes. Stereo Matching find corresponding pixels between stereo image pairs (a target image and a reference image).After finding the corresponding pixels, the disparity of a pixel is determined. Disparity describes the difference in the location of the two corresponding pixels in the stereo images when they are superimposed. The disparity can be converted to the depth value if the focal length of the camera and the distance between cameras are known [2]. However, finding the corresponding pixels between stereo images is not straightforward. The complexity of the stereo matching problem depends on the complexity of the 3D world scene.

Previous studies [3]-[5]indicate that there are constraints and stereo matching algorithms that can help reduce the number of ambiguous matches, but several unsolved problems still exist in stereo matching such as occlusion, discontinuity and repetitive pattern. Approaches to the stereo matching problem can be broadly classified into two categories: the area-based stereo matching algorithms and the feature-based stereo matching algorithms.

Area based techniques [6],[7] attempt to find corresponding pixels by matching image intensities, usually over the windows of fixed sizes in the two images. A very small window will not capture enough image structure, may be too noise sensitive or may produce many ambiguous matches, while a very large window makes matching less sensitive to noise, but results in very smooth disparity map. Hence, the size of the matching window is properly selected to avoid poor depth estimations. These techniques are easier to implement and generate dense disparity map, which is used for reconstructing surfaces. However, due to foreshortening effects and illumination changes, they are not suitable for matching stereo images captured from very different viewpoints. Moreover, these algorithms assume that all pixels in a finite matching window have similar disparities. This assumption may be false for pixels near depth discontinuities or edges. Hence, these algorithms generate unreliable disparity maps. These techniques provide good quality results in textured regions and ambiguous matches at occluded areas, texture less regions and edges.

Feature-based techniques [8]-[11] attempt to find corresponding pixels by matching features such as texture, segmentation, edge and corner in the two images. Because only matching pixels obtained from the extracted object features are used, these methods generate a sparse disparity map. These methods are faster and more accurate than area based matching techniques. The accuracy of feature based matching techniques depends on the number of reliable features detected. These algorithms generate reliable depth maps because they are comparatively insensitive to illumination changes, and are proper for applications such as visual navigation.

Al in 2012[9] presented a segment based stereo matching algorithm. He proposed a hybrid method that combines the advantages of two image segmentation algorithms: Belief Propagation and Mean Shift Segmentation algorithms. The Mean Shift algorithm is fast and Belief Propagation is very accurate segmentation. The proposed hybrid approach is then integrated with the Sum-of-Absolute Differences (SAD) stereo matching algorithm to obtain the disparity estimate of each pixel. This algorithm is speedy and generates highly accurate disparity. This technique has attracted attention due to its good performance on handling boundaries and texture less regions. Although this technique is interesting, it is sensitive to different image conditions i.e. rotation, occlusion, blurring, illumination, clutter, etc. and hence, failed to generate a reliable disparity map in such conditions.

Benedict et al. in 2014 [10] proposed an efficient and robust stereo matching algorithm that can work well with different image conditions. He used SURF (Speeded-Up Robust Features) and SIFTS (Scale Invariant Feature Transform) algorithms for extracting and matching corresponding points. The experimental results show that the SURF [11] [12] detector outperformed the SIFT [8] detector in terms of robustness, accuracy and computation time. It has been suggested that the SURF detector is be used for matching process and this seems to be a useful approach. However, a key problem with much of the literature in relation to SURF algorithm is that it produces a sparse disparity map, which is inadequate for reconstructing surfaces.

In this context we tried to analyze and implement a technique that produces a dense disparity map and that too with different image conditions. This proposed method incorporates open SURF [13] algorithm to deal with the different image conditions. The open SURF algorithm results are combined with the local window-based stereo matching method to obtain the disparity value of each pixel i.e., dense disparity map. The proposed approach achieves good result without using any image segmentation algorithms. Results show that our proposed HSAD-Hessian-open SURF algorithm improves the performance.

The remainder of the paper is organized as follows: A new methodology is presented in section II. Section III discusses the results. Section IV shows performance analysis which illustrates how the proposed algorithm works. Finally, section V concludes this paper.

2. Proposed work

In this section, we discuss the technique implemented in our proposed approach. The figure 1 shows the framework of Kamencay's approach using HSAD. He combined SAD algorithm with hybrid segmentation. The method of our proposed work is essentially the same as that used by Kamencay's [9] with some changes the following changes were made:

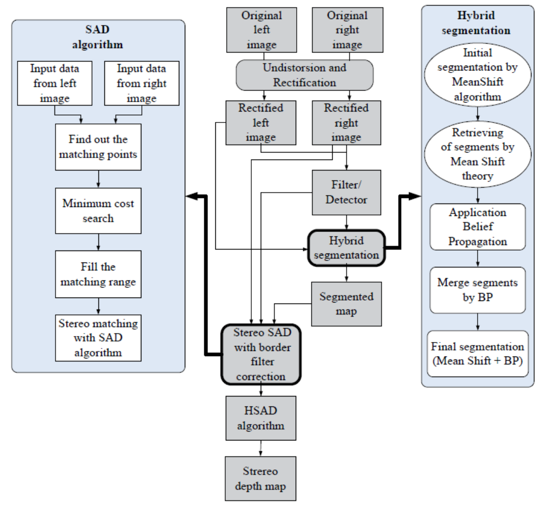

Fig. 1. Framework of Kamencay's approach

1. The hybrid segmentation (Mean Shift Segmentation and Belief Propagation) is replaced by a feature detection (open SURF) for matching corresponding points in the two images of the same scene. Although segment based techniques are good at handling texture less regions and boundaries, but they fail to find unambiguous corresponding points under different image conditions such as rotation, scale, illumination changes, occlusion, clutter, etc. Hence, they produce unreliable disparity maps. In an attempt to produce reliable disparity maps under different conditions, we use open SURF matching algorithm in our approach. Speeded-Up Robust Feature (SURF) [12] is a widely used feature detection and extraction algorithm in computer vision applications such as object recognition and 3D image reconstruction. A number of studies have found that SURF has a good stability against rotation, scale, occlusion and illumination changes. Moreover, SURF is more robust and faster than other existing feature detection algorithms such as SIFT. Among the different approaches of SURF, Open SURF [13] is an open source, clean and union structure implementation of SURF library. Open SURF algorithm combines the 4 steps for feature detection and matching as follows: Firstly, it detects features (depth points) that are likely to be found in the different images of the same 3D world scene. These depth points should be rotation and scale invariant. The corners (intersection of edges), blobs (sharp change in intensity), T-Junction, etc constitute important depth points and mostly searched in multiple scales. Open SURF algorithm uses box filters for depth point detection. Secondly, it finds the correct orientation of every depth point so that if the image is rotated then according to that orientation, both images are aligned in regard to that single depth point. Thirdly, it computes a descriptor or feature vector for every depth point. This feature vector represents how the neighborhood of the depth points look like after orientation in the correct scale. Finally, the feature vectors

are matched between stereo images. The matching is based only on a distance between the feature vectors i.e. Euclidean distance, not on depth point locations. The smallest Euclidean distance between the same depth points is considered as a good match.

2. Kamencay's [9] use of Mean Shift Segmentation (MSS)algorithm is fully justified. He applied MSS algorithm to filter image that is very useful for noise removing, smoothing and image segmentation. Hence, it is crucial to compensate for the MSS algorithm as our approach is not applying any image segmentation methods. In this context, the open SURF algorithm employ Hessian matrix for depth point detection because of its good performance in accuracy. The Hessian matrix facilitates the implementation of the concept of integral image which in turn increases the speed of detection process. More precisely, the open SURF algorithm detects features with sharp change in intensities at location where the determinant of Hessian Matrix is maximum. One significant reason to implement open SURF technique is that it gives complementary information about the depth point that cannot be obtained from corner or edge detectors. Moreover, the improvement on speed achieved with open SURF algorithm is a very desirable factor for real time application systems.

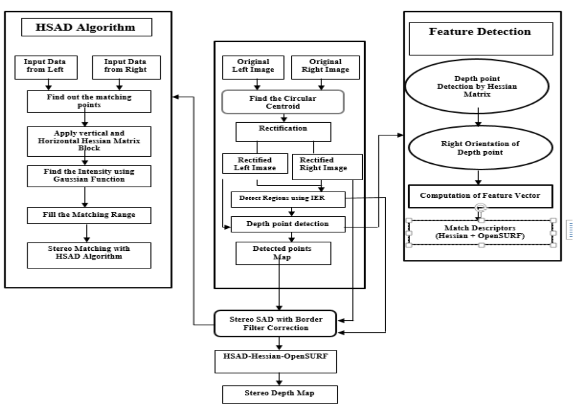

Fig. 2. Framework of proposed algorithm using HSAD-Hessian-Open SURF (combines Open SURF algorithm and SAD algorithm)

3. Implementation of proposed algorithm

In this section, we describe the implementation of our new approach.

1) *Read stereo image pair:* In this step, we select the stereo image pair from Middlebury [3] data set as shown in figure 3. We read the reference (right) image first and then the target image (left). We found that the left portion is slightly cropped in the left image and the right portion in the right image.
2) *Detects depth points in both images:* This step detects the 10 strongest depth points that are likely to be found in both images using feature matching algorithm (open SURF) as shown in figure 4. These points are rotation and scale invariant. The figure 4 shows that the centric of the depth point stops at the position in the target image which is cropped in the reference image to speed-up the detection.
3) *Orientation of depth points:* This step finds the right orientation of detected depth point, so that if the image is rotated then according to that orientation both images are aligned to that single depth point.
4) *Extract feature vector and match vectors:* This step first calculates the feature vectors at the depth point in both images and then matches these vectors between both images as shown in figure 5. These vectors contain the information of how the neighborhood of depth points looks like after correct orientation. These vectors are insensitive to noise, geometric and photometric deformations, and detection errors. The small Euclidean distance between the same points constitute a good match.
5) *Compute Disparity map:* The open SURF matching algorithm improves accuracy of the disparity calculation. However, the calculated disparity map is sparse because only the detected depth points are matched. In this step the results of the Open SURF algorithm are integrated with the SAD stereo matching algorithm to calculate the disparity value of each pixel in the image i.e. to produce dense disparity map. The final disparity map is shown in figure 6. The disparity map is generated by only searching block to the right direction. The block size of 7x7 pixels is used. Interpolation is performed between the nearest matching window and its neighbors to fine-tune the disparity value to a sub pixel location.

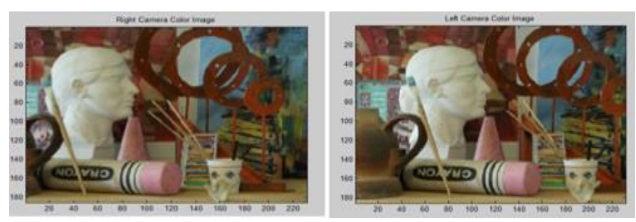

Fig. 3. Right and left views of the ART stereo image pair collected from middlebury dataset [3]

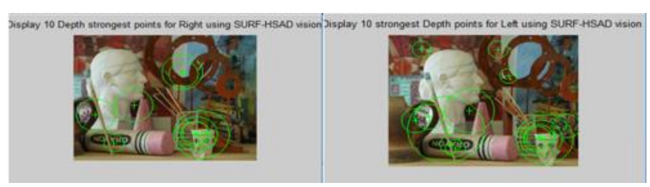

Fig. 4. Display of strong depth points in right and left stereo images using openSURF algorithm

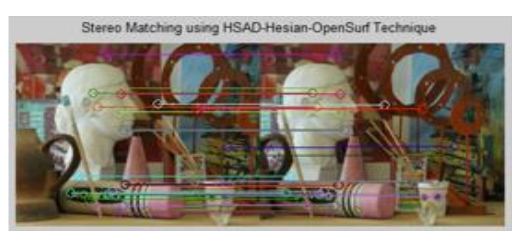

Fig. 5. Stereo matching using HSAD-Hessian-openSURF technique

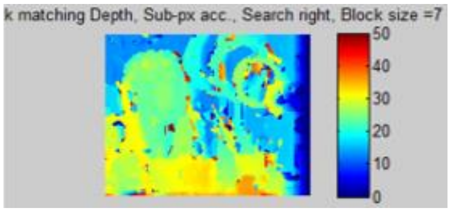

Fig. 6. Final disparity map using HSAD-Hessian-openSURF technique

4. Results

In this section, the results of proposed method are evaluated on a Middlebury dataset [3]. Results of our proposed method is compared with existing method are given in graphs. Our proposed HSAD-Hessian-Open SURF technique can provide depth information under different image conditions (i.e. illumination, rotation, blurring, scale, clutter, etc.). The proposed technique outperforms the existing HSAD technique, 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 on 2.19 GHz Intel i5-5200U CPU. This shows that the proposed method is suitable for real time applications. The simulation is performed using MATLAB software (R2014a).

5. Performance evaluation

The performance evaluation of proposed and existing methods is presented in this section, to find the best among them that would allow us to get depth information even under different image conditions.

The parameter which is widely used for analyzing matches is based on precision and recall measurement. Recall represents the ability to find all the unambiguous matches. Precision expresses the ability to get unambiguous matches when the number of matches obtained varies. The figure 7 represents the precision-recall curve for both techniques for ART stereo image pair in the data set. A precision-recall curve [10] representing high precision value with a low recall value shows that we have obtained unambiguous matches, but many others have been missed. A high recall value with a low precision value shows that we have obtained mostly unambiguous matches but there are also lots of ambiguous matches. For this very reason, it is desirable to find a technique that employs descriptor that get high values of both precision and recall simultaneously, thus having values located at the upper-right corner in the precision versus recall curve. The proposed algorithm gives better results in terms of precision and recall than existing.

F-measure combines both precision and recall values into a single value, so that we just need to select the unambiguous depth point in the image which has the highest F-measure.

$$F = \frac{2PR}{P+R}.100\%$$

For each pair of precision and recall we calculated F-measure and then picked the depth point which has the highest F-measure's-measure values precision and recall the same, but sometime the recall is more important than precision. (For example. We don't mind having a lot of points ambiguously detected as depth points if we know that all the points which are depth points are detected). F-measure reaches its best value at 1 and worst value at 0. The figure 8 shows F-measure – Recall curve for both techniques for ART stereo image pair in the data set and proposed technique gives better results.

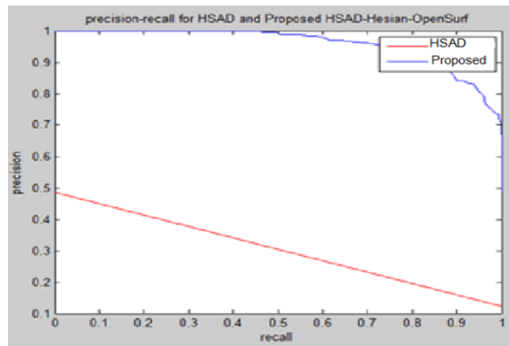

Fig. 7. The Recall-Precision curve under different image conditions

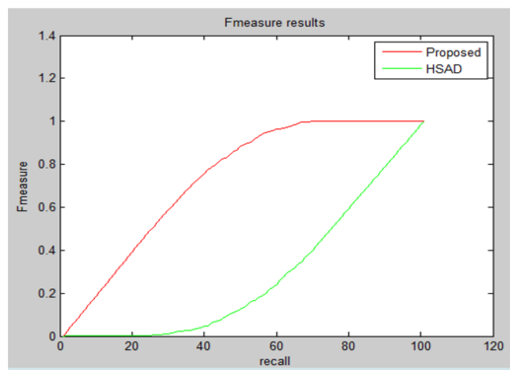

Fig. 8. F-measure- Recall curve under different image conditions

The figure 9 shows the accuracy - k-point curve for both techniques for ART stereo image pair in the data set. The figure represents the accuracy of the algorithms to detect the same number of depth points unambiguously under different image conditions.

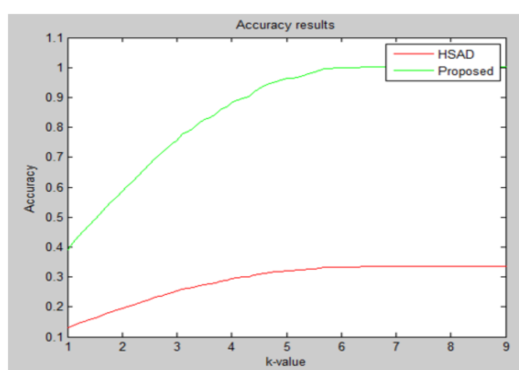

Fig. 9. Accuracy- k-value curve under different image conditions

It is clear from the figure that the proposed HSAD-Hessian-

Open SURF technique is 97% accurate, whereas the HSAD algorithm is only 30 % accurate.

6. Conclusion

In this paper, an efficient, flexible and robust HSAD-Hessian-open SURF technique has been introduced to overcome the limitations of hybrid segmentation algorithm. The proposed approach firstly detects and matches depth points using open SURF algorithm, and then combines the obtained results with SAD stereo matching algorithm to produce the dense disparity map. The proposed technique can provide depth information under different image conditions (i.e. illumination, rotation, blurring, scale, clutter etc.). The Open SURF algorithm utilizes Hessian Matrix to detect and match depth points. Hessian Matrix minimizes the filtering process and thus, speedup the search for corresponding points. The results show that our technique outperforms the hybrid segmentation algorithm, both in speed and accuracy. The proposed technique is proper to real-time application systems like 3D image reconstruction, robot navigation, mapping, etc.

References

[1] Trucco, E., & Verri, A., Introductory Techniques for 3-D Computer Vision. New Jersey, Prentice Hall, 1998.
[2] Belhumeur, P.N., A Bayesian Approach to Binocular Stereopsis. In Int'l J. Computer Vision, vol. 19, no. 3, pp. 237-260, 1996.
[3] Scharstein, D., & Szeliski, R., The Middlebury Stereo Vision Page, 2002. http://vision Middlebury.edu/stereo/.
[4] Scharstein, D., & Szeliski, R., "A taxanomy and evaluation of dense two frame stereo correspondence algorithms," in International Journal of Computer Vision, vol.47, no. 1, pp.7-42, 2002.
[5] Mattoccia, S. (2011). Stereo vision: algorithms and applications, www.vision.deis. unibo.it/
[6] F. Tombari, S. Mattoccia and L. Di Stefano, "Stereo for robots: Quantitative evaluation of efficient and low-memory dense stereo algorithms," *2010 11th International Conference on Control Automation Robotics & Vision*, Singapore, 2010, pp. 1231-1238.
[7] X. Sun, X. Mei, S. Jiao, M. Zhou and H. Wang, "Stereo Matching with Reliable Disparity Propagation," *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, Hangzhou, 2011, pp. 132-139.
[8] Juan, L., & Gwun, O., A Comparison of SIFT, PCA-SIFT and SURF. International Journal of Image Processing, Vol. 3, Issue 4, 2009.
[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, 2012.
[10] Benseddik, H., Djekoune, O., & Belhocine, M., SIFT and SURF Performance Evaluation for Mobile Robot-Monocular Visual Odometry. In Journal of Image and Graphics, Volume 2, No.1. 2014.
[11] Bay, H., Tuytelaars, T., & Gool, L.V., Surf: Speeded up robust features. In European Conference on Computer Vision, 1:404-417, 2006.
[12] Lowe, D. G. (1999). Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision.2, pp.1150-1157, 1999.
[13] Evans, C. OpenSURF - Open Source SURF feature extraction library. http://code.google.com/p/opensurf1/