ISSN (Print) : 2320 – 9798
ISSN (Online): 2320 – 9801

International Journal of Innovative Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2013

Advancement in Depth Estimation for Stereo Image Pair

Shabnam Hussain[1], Ravindra Modi[2]

Master of Technology, Dept. of Computer Engineering, U V Patel College of Engineering, Kherva, Gujarat, India[1]

Assistant Professor, Dept. of Information Technology, U V Patel College of Engineering, Kherva, Gujarat, India[2]

ABSTRACT: Many researchers invented ideas to compute depth estimation. In Hybrid Technique for Stereoscopic Depth Estimation there were some drawbacks associated with the optical flow estimation when it was done with novel disparity compensation step that the edges between the objects in disparity maps were not preserved very well which was caused in filtering step of optical flow analysis, related with desired smoothness. This research work presents an advance technique for the disparity map estimation from a stereo pair of images. The original concepts of proposal are: use of disparity estimation, a new hierarchical shape-adaptive block matching, optical flow analysis, novel occlusion detection and disparity extrapolation schemes. It is fairly insensitive to parameter variations, and it indicates its excellent robustness under noise. This method gives significantly smaller angular errors than previous techniques for optical flow estimation. The main advantages of this proposal are: low computational complexity, subpixel accuracy of depth estimation, operation across flat or untextured regions and good detection of occluded regions. The results of simulation are shown and compared different quality parameter of it's by applying on various images.

KEYWORDS: Disparity, Occlusion, Optical flow.

I. INTRODUCTION

The distance between the surface of scene objects from a viewpoint. The depth estimation world is a quite complex research field, where many techniques and setups have been proposed. The set of algorithms which solve the depth map estimation problem deals with many different mathematical concepts which should be briefly explained for a minimum overall comprehension of the matter.Disparity map refers to the apparent pixel difference or motion between a pair of stereo image. To experience this, try closing one of your eyes and then rapidly close it while opening the other. Objects that are close to you will appear to jump a significant distance while objects further away will move very little. That motion is the disparity. In a pair of images derived from stereo cameras, you can measure the apparent motion in pixels for every point and make an intensity image out of the measurements.

The analysis of sequence of images, used to approximate motion, is a core area of computer vision. The 3-D velocity vector of objects, projected onto the image plane, is known as the image flow field. This could be considered the ideal, the actual movement of objects that we expect to see. Unfortunately, we are often working in the reverse direction: given a sequence of images, we want to determine the movement of the objects. This leads to an approximation called the optical flow field [1] which corresponds to the movement of pixels in an image, or what we actually see. It is easy to imagine situations where sequence of image do not correspond with the actual object movement.

II. BACKGROUND STUDY

The area of stereoscopic depth processing has been and still is one of the most actively researched fields in computer vision. Today there is a vast amount of algorithms that use very diverse principles for finding the pixel correspondences that are necessary to calculate 3-D depth. The major applications are 3-D scene vision for mobile platforms and 3-D reconstruction of object surfaces. While mobile platforms usually employ a parallel stereo camera setup, the cameras used for 3-D reconstruction are often in a general position. It has been observed in [2] that all stereo matching algorithms typically consist of four main steps: Matching cost computation, Cost aggregation, Disparity computation, Disparity refinement

In the first step a matching cost between all potentially corresponding pixels is calculated. Common pixel-based matching costs include squared difference, absolute difference, binary matching costs or gradient- based measures [2]. The corresponding pixels in rectified stereo images are located on the same image line. This enables to store the matching costs in a condensed data structure called Disparity Space Image (DSI). The DSI is a three-dimensional structure whose entries are indexed by x-position, y-position and disparity.

After the DSI has been generated and improved by aggregation, the pixel correspondences have to be determined. The most simple form is a maximum selection, i.e. for each pixel in one image the best matching pixel in the other image is selected. More complex selection mechanisms incorporate smoothness constraints for neighboring pixels to reduce ambiguities and increase the robustness to noise. The last step of a stereo computation is a refinement of the calculated disparities. These encompass interpolation of disparities to sub-pixel accuracy or a left-right consistency check for detecting occlusions.

In the following, some of the currently most prevalent stereo computation algorithms are introduced and analyzed with respect to their real-time capability and quality. Usually, the main idea of these algorithms is to focus on one of the first three steps of the stereo computation, while being quite unconstrained with respect to the other steps. The analysis will reveal that only a few approaches solely scale linearly with the number of pixels n and number of disparities d, which is a prerequisite for real-time stereo processing.

A. *Global Stereo Matching*

Global stereo matching approaches usually turn the stereo matching problem into an energy-minimization problem. The goal is to find a disparity function D which minimizes a global energy function

$$E(D) = E_{data}(D) + \lambda E_{smoot\ h}(D) \qquad (1)$$

where E_{data} (D) reflects the matching quality of the two stereo images given the disparity map D. Hence, E_{data} (D) constitutes the aggregation of the matching costs at a pixel-level. The second term $E_{smoot\ h}$(D) makes the energy function favor smooth transitions between the disparity of neighboring pixels (continuity constraint). In general the smoothness term adds some increasing penalty for increasing difference of disparity that may also depend on the intensity or color difference of the pixels.

Dynamic Programming

As mentioned above, finding the global optimum of the global energy (1) is NP-hard. One way to tackle this problem is to solve the optimization for each image line independently. By doing so the global 2-D optimization is split up into multiple 1-D optimizations. These 1-D optimizations can be done efficiently in polynomial time using Dynamic Programming (DP) [3]. The optimization scheme follows Bellman's principle of optimality and constitutes a shortest-path search in a graph, where the nodes of the graph are the elements of the DSI and the path lengths are the costs defined by the energy function (1).

Dynamic programming is one of the techniques that are frequently used in real-time implementations [4] because it allows for an efficient improvement of the results from a correlation-based stereo. Unfortunately, DP has some drawbacks that limit its usage. Firstly, DP has an inherent demand for the ordering constraint, i.e. DP assumes that the ordering of pixels stays the same for corresponding image lines of the two stereo views. This may be violated by narrow foreground objects. Secondly, DP leads to the so-called streaking effect, which is a tendency of DP to produce fine streaks in the disparity map, in particular at depth discontinuities.

Graph Cut

Graph Cut (GC) approaches for global stereo matching [5] have become very popular for the optimization of the energy function (3.4). In its true nature GC actually is a labeling technique that tries to find an optimal label assignment f based on the energy function:

$$E(f) = E_{data}(f) + \lambda E_{smoot\ h}(f) \qquad (2)$$

In case of stereo matching the labels f become disparities D. An efficient variant of GC was proposed in [6]. The main difference of this approach compared to other label assignment approaches is that the algorithm allows for a simultaneous change of the labels of a large number of pixels. These changes are done by so called α-β-swaps or α-expansions. The first one is able to exchange two labels α and β and the latter one is able to change any label to the label α. Using one of these moves in an iterative fashion leads to a locally optimal solution. When using the α expansions the found local optimum is within a known factor of the global optimum.

After this operation all pixel vertices are only connected to one of the two label vertices which indicates the new label for each pixel. In general GC produces spatially dense and precise depth maps. The downside is the large runtime compared to standard correlation-based stereo.

Belief Propagation

Similar to GC, Belief Propagation (BP) solves low-vision problems like image restoration or segmentation by reinterpreting these as a label assignment problem on a graph structure. The label assignment problem constitutes a minimization of the energy function (3.5). For stereo calculation these labels turn into disparities. The difference between BP and GC is the way the graph structure is created and used. While GC tries to map the problem to a minimum cut, BP uses a more direct mapping. For BP each pixel becomes a vertex and edges are usually drawn to the direct neighbours, i.e. in the form of a four-connected image grid. The algorithm optimizes the energy function (2) by starting with a random label assignment to the nodes and then improves the assignment by passing messages around the graph. The messages are vectors of a size given by the number of labels. They encode probability distributions of the label assignment of neighbouring pixels. One major drawback of BP is the large amount of data that has to be stored. For very large images even the main memory might not have sufficient space which limits the application of BP in stereo vision to smaller images.

B. Semi-Global Matching

As the name suggest Semi-GlobalMatching (SGM) [7] constitutes a compromise between global and local stereo correspondence matching. The motivation was to combine the efficiency of local methods with the accuracy of global optimization approaches. Like global matching approaches SGM uses a pixel-wise cost and a smoothness constraint for defining an energy E(D) that has to be optimized for the disparity image D.

$$E(D) = \sum_p C(p, D_p) + \sum_{q\epsilon N_p} P_1 \delta[|D_p - D_q| == 1] + \sum_{q\epsilon N_p} P_2 \delta[|D_p - D_q| > 1] \quad (3)$$

The energy (3.8) of SGM is split into three terms. The first one sums up the matching cost for the disparities of D. The second and third term express the smoothness by adding a small penalty P_1 for neighboring pixels that differ only by one disparity from the disparity of pixel p and a larger penalty P_2 for larger disparity differences. The idea behind the smaller penalty is to permit for the small changes that occur on slanted and curved surfaces. The penalty for larger changes restricts large disparity jumps to large intensity jumps in the image. A global optimization of D with respect to the energy E(D) is N_p complete. On the other hand, the minimization of such an energy function can be done in polynomial time if one performs the minimization along all rows individually, using Dynamic Programming (DP).

It was suggested in [7] to use at least eight paths but better 16 paths for having a good coverage of the 2-D image. It has been shown in [7] that the runtime of SGM is O(lnd), where l is the number of paths, n the number of pixels and d the number of disparities. As l is usually eight or 16 the runtime of SGM is much smaller than the runtime of global optimization approaches but still higher than for standard block-matching stereo. One disadvantage of SGM is the necessity of storing the full disparity distribution for all pixels which leads to a space complexity of O(nd). As it has been discussed for BP this can lead to a significant decrease in speed on modern computers and limits the application to smaller stereo images.

III. PROPOSED METHOD

A.Description of Algorithm

Read Stereo Image Pair

Here we read in the color stereo image pair and convert the images to gray scale for the matching process. Using color images may provide some improvement in accuracy, but it is more efficient to work with only one-channel images. For this we use the ImageDataTypeConverter and the ColorSpaceConverter System objects. Below we show the left camera image and a color composite of both images so that one can easily see the disparity between them.

Hierarchical shape-adaptive block matching

The block matching technique produces initial guess of disparity map for further processing. Hierarchical block matching iteratively estimates disparity map starting from decimated images of the lowest resolution and improves accuracy and resolution in subsequent steps. Moreover size of the blocks employed by this technique is adaptively selected. Next we perform basic block matching. For every pixel in the right image, we extract the 7-by-7-pixel block around it and search along the same row in the left image for the block that best matches it. Here we search in a range of ± 15 pixels around the pixel's location in the first image, and we use the sum of absolute differences (SAD) to compare the image regions. We need only search over columns and not over rows because the images are rectified. We use the TemplateMatcher System object to perform this block matching between each block and the region of interest. Re-running basic block matching, we achieve the result below where the contouring effects are mostly removed and the disparity estimates are correctly refined. This is especially evident along the edges.

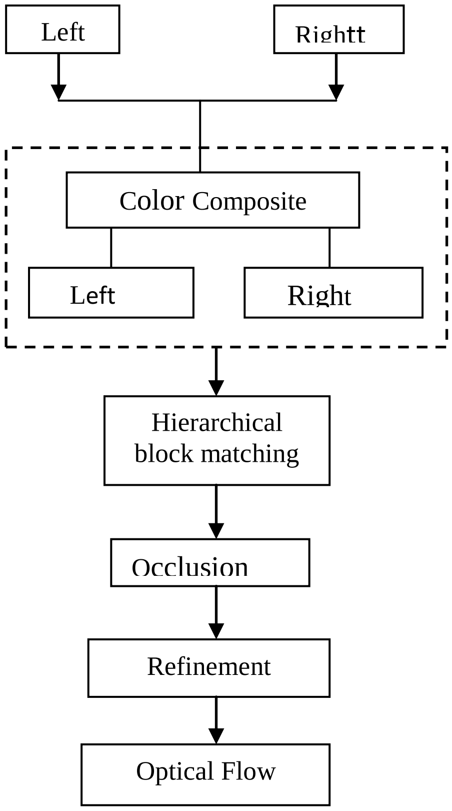


Figure 1: Proposed Method

Occlusion detection

Not all pixels in the left image have corresponding pixels in the right image (and vice versa) because of occlusion. Regions in disparity map that correspond to occluded regions contain artifacts. We propose a method that replaces these erroneous values in disparity map.

Our experiments revealed that straightforward use of NSAD (as pixels in occluded regions mismatch and imply relatively big values of NSAD) it is not efficient, because it is uncertain whether big NSAD values refer to occluded regions or result from existence of noise. Occlusion is detected with analysis of pixel usage produced by block-matching. The pixels that are unmatched by complimentary path are classified as occluded. In such an approach left path generates occlusion information for right disparity map and vice versa.

Dynamic Programming

As mentioned above, basic block matching creates a noisy disparity image. This can be improved by introducing a smoothness constraint. Basic block matching chooses the optimal disparity for each pixel based on its own cost function alone. Now we want to allow a pixel to have a disparity with possibly sub-optimal cost for it locally. This extra cost must be offset by increasing that pixel's agreement in disparity with its neighbours. In particular, we constrain each disparity estimate to lie with ± 3 values of its neighbours disparities, where its neighbours are the adjacent pixels along an image row. The problem of finding the optimal disparity estimates for a row of pixels now becomes one of finding the "optimal path" from one side of the image to the other. To find this optimal path, we use the underlying block matching metric as the cost function and constrain the disparities to only change by a certain amount between adjacent pixels. This is a problem that can be solved efficiently using the technique of dynamic programming[8]. It does nothing to smooth "between" rows, which is why a striation pattern appears. Despite these limitations, the result is significantly improved, with the noise nearly completely removed, and with many of the foreground objects being better reconstructed.

Optical flow

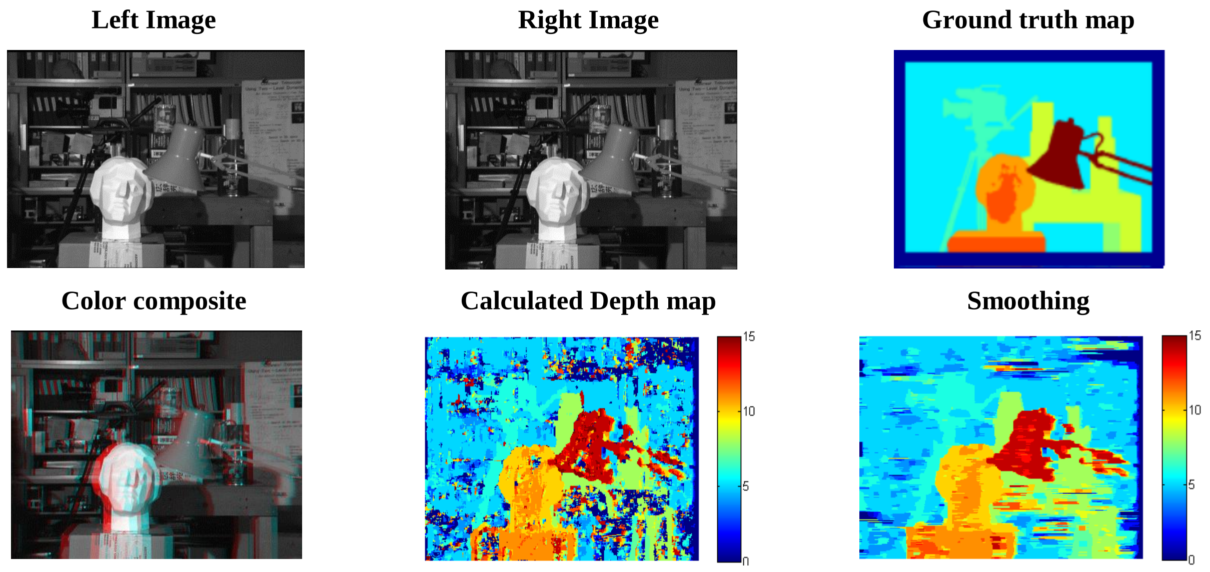
In computer vision, the Lucas–Kanade [9] method is a widely used differential method for optical flow estimation developed by Bruce D. Lucas and Takeo Kanade. It assumes that the flow is essentially constant in a local neighbourhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighbourhood, by the least squares criterion. By combining information from several nearby pixels, the Lucas–Kanade method can often resolve the inherent ambiguity of the optical flow equation. It is also less sensitive to image noise than point-wise methods.

The least-squares approach implicitly assumes that the errors in the image data have a Gaussian distribution with zero mean. If one expects the window to contain a certain percentage of "outliers" (grossly wrong data values, that do not follow the "ordinary" Gaussian error distribution), one may use statistical analysis to detect them, and reduce their weight accordingly. The Lucas–Kanade method can be used only when the image flow vector (V_x, V_y) between the two frames is small enough for the differential equation of the optical flow to hold, which is often less than the pixel spacing. When the flow vector may exceed this limit, such as in stereo matching or warped document registration, the Lucas–Kanade method may still be used to refine some coarse estimate of the same, obtained by other means; for example, by extrapolating the flow vectors computed for previous frames, or by running the Lucas-Kanade algorithm on reduced-scale versions of the images. Indeed, the latter method is the basis of the popular Kanade-Lucas-Tomasi (KLT) feature matching algorithm.

IV. SIMULATION AND RESULTS

A.Depth Estimation by Block matching

Block matching is applied on image for estimating Depth, result is shown in below figure. Depth map of Tsukuba and Cones after smoothing by dynamic programming can be compared to its ground truth map. Edges and objects are clearly visible and also the presented algorithm is capable to reveal background information. Disparity Range: 0-18, Window Size: 3x3, The hotter the color, the closer it is; the cooler the color the farther it is.



Left Image

Right Image

Ground truth map

Color composite

Calculated Depth map

Smoothing

Figure2: Depth Estimation of Tsukuba by Block Matching

TABLE I : Comparison of Bad pixels in Tsukuba with other algorithms

Tsukuba	Bad pixels	NBP-SAD	NBP-SSD	PSNR
Our proposal	1.0%	2.9	9.9	29.61
Adapting BP [10]	1.58%	2.50	24.95	31.51
AdaptOvrSegBP [11]	1.22%	4.16	25.95	29.61
PlaneFitBP [12]	0.89%	4.10	25.46	35.32
DoubleBP [13]	5.25%	5.15	33.75	33.19
SubPixDoubleBP [14]	1.10%	5.28	34.78	29.62
SSD +MF [15]	0.96%	5.25	35.71	34.87
Ground truth	0.00%	-	-	35.31

Table shows results attained by our technique compared with quality measures mentioned above. Among other techniques, proposed technique has the best bad-pixel ratio, but average NBP-SAD value. It means that it produces less bad-pixels, but with relatively some errors versus ground truth. The values of PSNR and NBP-SSD show that devised technique is competitive to other algorithms and referring to them, places at about the middle of the ranking.

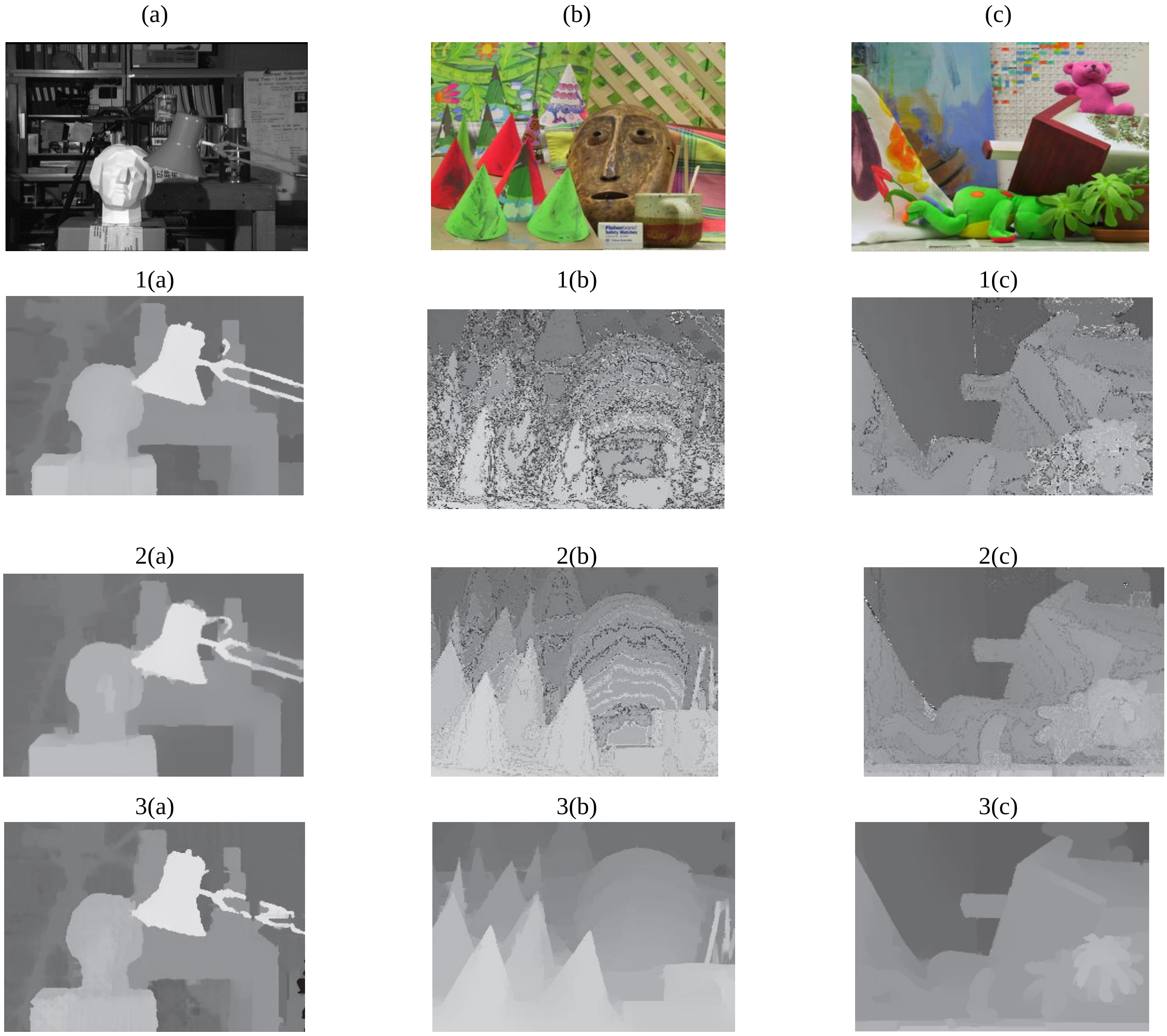


Figure 3: Images (a) Tuskuba, (b) Cones, (c) Teddy, 1(a), 1(b), 1(c) are the resulting depth maps by Cost Aggr., 2(a), 2(b), 2(c) are resulting depth maps by Semi-global methods and 3(a), 3(b) and 3(c) are depth maps from our proposed methods.

Comparison in RMSE:

Root Mean Square Error (RMSE) is calculated by

$$\text{RMSE}=\left(\frac{1}{N}\sum_{(x,y)}\left|d_c(x,y)-d_t(x,y)\right|^2\right)^{1/2}$$

where, N is the total number of pixels in the image, $d_c$is the computed disparity map, d_t is the ground truth map.

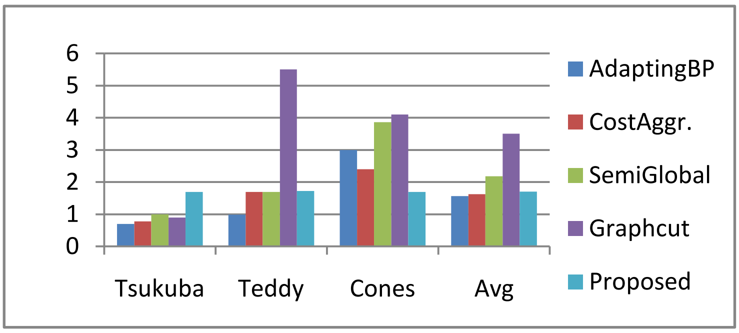


Figure 4: Graphical representation of RMSE comparison.

Root Mean Square Error (RMSE) to the ground truth and compared with the results from other algorithms. Figure shows the result where Adapting-BP is a method using belief propagation which is ranked at the top in the BPP(Bad Pixel Percentage) test and Graph-cut is an alpha-expansion method with occlusion handling which is ranked at 42th. All algorithms compared in this test are ranked higher than the proposed algorithm in the BPP test. We can see that the proposed method shows much better performance compared with Semi-global matching and Graph-cut, and compete with Adapting-BP and Cost-Aggregation methods in the RMSE test.

V. CONCLUSION

The block matching algorithm is implemented in this project to achieve correspondence between two stereo images. In the process of block matching, a block of fixed size is centered on each extracted feature point in an image, and then the intensity difference is calculated between any block from the left image and any block from the right image and a modified smoothing algorithm is developed to improve the depth estimation performance from the extracted viewpoint images which helps in achievement of accurate depth measurement result and acceptable depth maps from the images.

REFERENCES

[1]. B.K.P. Horn, B.G. Schunck, "Determining optical flow", Massachusetts Instiue of Technology, 1980.
[2]. Scharstein, D. and Szeliski, R, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision, 47(1-3), pp.7–42, 2002.
[3]. Md Cormen, T. H., Rivest, R. L., Leiserson, C. E., and Stein, C. Introduction to Algorithms. MIT Press, 3rd edition, pp.23-50, 2009.
[4]. Salmen, J., Schlipsing, M., Edelbrunner, J., Hegemann, S., and Luke, S. Real-time stereo vision: Making more out of dynamic programming. In Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns (CAIP), pp. 1096–1103, 2009.
[5]. Kolmogorov, V. and Zabih, R. Computing visual correspondence with occlusions using graph cuts. In Proceedings ot the International Conference on Computer Vision (ICCV), pp. 508–515, 2001.
[6]. Boykov, Y., Veksler, O., and Zabih, R. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(11):1222–1239. 2001.
[7]. Hirschmuller, H. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 807–814.2005
[8]. Park, CS; Park, HW. "A robust stereo disparity estimation using adaptive window search and dynamic programming search" 2000.
[9]. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", Proceedings of Imaging Understanding Workshop, pp. 121—130, 1981.
[10]. Anonymous, "Stereo reconstruction with mixed pixels using adaptive over-segmentation", CVPR, 2008.
[11]. Q. Yang, R. Yang, J. Davis, D. Nister. "Spatial-depth super resolution for range images", CVPR ,2007.
[12]. Q. Yang, L. Wang, R. Yang, "Stereo matching with colorweighted correlation, hierarchical belief propagation and occlusion handling", CVPR 2006.
[13]. D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense twoframe stereo correspondence algorithms", IJCV, 2002.
[14]. A. Klaus, M. Sormann, K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", International Conference on Pattern Recognition, 2006.
[15]. O. Stankiewicz, K. Wegner, "Depth Map Estimation Software version 2", MPEG 2008/M15338, Archamps, France, April 2008.

BIOGRAPHY

SHABNAM HUSSAIN received the B.E. degree in Information Technology from the Gautam Buddha University, in 2011, the final year of M.TECH pursuing in Information Technology from the University of Ganpat, Kherva, Mehsana, Gujarat, India. His research areas include Image processing.