International Journal on Emerging Technologies (Special Issue on ICRIET-2016) **7**(2): 127-132(2016)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Local Stereo Parametric Methodsof Disparity Computation using GPU

Priya Charles* and Dr. A. V Patil**
**VIIT research scholar, Electronics & Telecommunication department,*
VIIT,Pune,India-411044
***Principal, DYPIEMR, Pune, India-411044*

(Corresponding Author: Priya Charles)
(Received 16 September, 2016 Accepted 19 October, 2016)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Vision is the most important sense for humans and because of this human vision system only we are able to see the 3D world around us with great clarity and are able to find out depth of each and every object. Many Active and Passive depth estimation techniques have been proposed which are capable of estimating depth of real world scene among which one of the passive method, stereo vision has been proven to provide remarkable results. We have used stereo vision technique to estimate depth for a given real world scene. We have done calibration and used non rectified as well as rectified stereo images and then algorithms such as SAD, SSD, NCC, SAD by Derivatives are used for estimating reliable and accurate correspondence match for stereoscopic image pairs. We have used window and aggregation approach to improve the accuracy of disparity map and triangulation method is used to compute depth from disparity space image matrix .The algorithms are implemented in multicore processors using CUDA 7.5 tool kit in windows operating system. CUDA is used to calculate time taken for disparity computation. The output of depth estimated is in form of a matrix which we process in parallel manner using multiprocessor method. Results of the data base images have been compared with online available Middlebury datasets and their values.

Keywords: stereo image, matching cost computation, CUDA, disparity

I. INTRODUCTION

We see the 3D world with our eyes and also know that human vision system is a very important system that enables us to compute and visualize the things that are in scope of our eye with clarity and at accurate distances. We see and analyze the actual distances of objects with our naked eye and also calculate comparative distances between objects. This system helps in our day to day lifestyle and work. But when we click images of a 3Dworld system this third dimension of depth gets lost hence we are unable to compute the actual distance of the objects from the point where sensors were placed and also the relative distance between objects. With advancement in the computer vision and high speed computing units we are now able to compute this third dimension which was lost earlier and this method of computing the lost depth is known as Depth Estimation as shown in figure 1. A stereo correspondence algorithm matches pixels of one image (reference) to pixels of the other image (target) and returns the corresponding vertical displacement as the reference pixel's disparity, which is proportional to its depth.

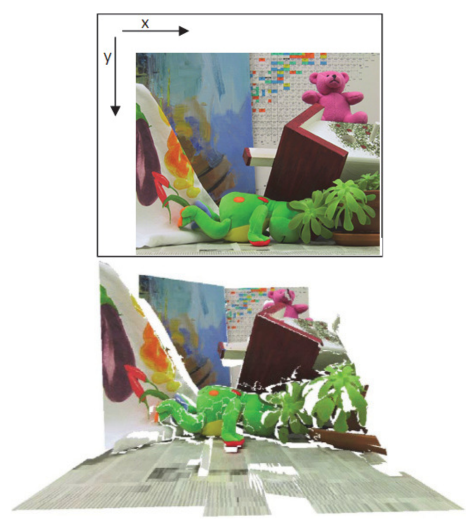

Fig. 1. 2D projection of a scene and its 3D reconstruction [2].

In traditional stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene, in a manner similar to human binocular vision. By comparing these two images, the relative depth information can be obtained, in the form of disparities, which are inversely proportional to the differences in distance to the objects. To compare the images, the two views must be superimposed in a stereoscopic device, the image from the right camera being shown to the observer's right eye and from the left one to the left eye. In real camera systems however, several pre-processing steps are required. The image must first be removed of distortions, such as barrel distortion to ensure that the observed image is purely projectional. The image must be projected back to a common plane to allow comparison of the image pairs, known as image rectification. The displacement of relative features is measured to calculate a disparity map. Optionally, the disparity as observed by the common projection is converted back to the height map by inversion. Utilizing the correct proportionality constant, the height map can be calibrated to provide exact distances. Thus, stereo vision is able to retrieve the third dimension of scenery and therefore, its importance is obvious in issues such as traversability estimation, robot navigation, simultaneous localization and mapping (SLAM), as well as in many other aspects of production, security, defense, exploration and entertainment.

The remainder of this brief is organized as follows. Section II provides a background of various stereo matching techniques. Section III presents the theory behind various local stereo matching techniques. Sections IV presents quality metric theory for evaluation and its mathematical formulae and section V provides results for multiple stereo matching techniques. Section VI draws the conclusion.

II. BACKGROUND

In computer vision a set of algorithms are implemented to process images in a complex way, where human visual system is an important source of inspiration. Thus, many algorithms are trying to achieve human capabilities. There are many approaches to obtain the distance of a point (or a set of them). Generally we can divide all the methods to electronically measure the distance as active and passive. Active methods put some energy in the scene, projecting it in order to, in some way, illuminate the space, and processing, passively, the reflected energy[3] and [4]. These methods were proposed before the passive ones, because of one main reason: the micro-processing was not even invented. These methods present the main disadvantage, regarding the passive ones, in the energy needed. However, their accuracy use to be much higher, and some of them are used to obtain the ground truth. Active methods are those which do not require any energy to be projected on screen instead they use some sensors to find out depth. Active methods are further classified into light based and ultrasound based methods. Light based methods include incandescent light and time of flight based techniques. Passive methods capture the images with image sensors, being the problem solved in a computational way.

Passive methods are classified into mainly two groups: monocular and multiview.Monocular solutions for the depth estimation uses single image (or a video sequence of them) to obtain the depth map. The main limitation of this approach is the intrinsic limitation of the depth characteristics lost during the projection of the scene into the image plane. An advantage of this approach uses to be the relatively low amount of operations needed to process one single image, instead of two or more [5], [6], [7] and [8]. Multiview solutions for depth estimation uses two or more images to compute the depth map. Stereo vision is a particular case of this set, using two images whereas multi-view uses more than two images. Research area that could also be benefited by use of multiple camera arrays is so called co-operative stereo vision i.e multiple stereo cameras being considered to improve depth [9] and [10]. Here we will focus only on stereo vision techniques.

Stereo methods can be further classified into those producing sparse and dense outputs. Robotic application such as traversability, obstacle avoidance demand dense outputs hence we have to go for stereo vision methods that produce dense outputs [16]. Dense matching algorithms are basically of two types:

1) Local method: these are the methods that are based on area calculation. These methods trade accuracy for speed [18].

2) Global methods: these are the methods that are based on energy calculation. These methods trade speed for accuracy i.e more accurate results are obtained with higher computation time. Here we have only focused on local methods. Various local matching methods exist e.g Sum of Absolute difference (SAD)[11], Sum of Squared differences(SSD)[12], Normalized Cross Correlation(NCC) [13] e.t.c. These methods are classified as parametric methods because they depend directly on intensity values.

A multiprocessor is a computer system having two or more processing units each sharing the main memory and peripherals in order to simultaneously process programs.

III. STEREO MATCHING TECHNIQUES

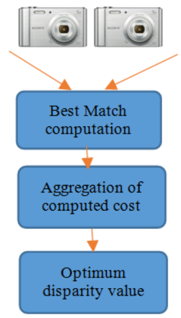

Fig. 2. Steps in computing disparity.

The matching cost function is a measure that quantitatively expresses how much dissimilar as shown in Figure 3(or equivalently similar) two image pixels are [14] and [15].

Algorithm to compute disparity:
Start
Step 1: define min and max disparity values, win size
Step 2: read input images (left and right)
Step 3: convert if colour to grayscale
Step 4: for disp in range (min disp to max disp)
Step 5: for pixel in range (-win to +win)
Step 6: compute sad, ssd, ncc, sad by derivatives
Step 7: if sad<prevbest then prevbest=sad
Step 8: goto step 5
Step 9: goto step 4
Step 10: create disparity map matrix
Step 11: multiply by scale factor to convert to grayscale
End

SAD (Sum of Absolute Differences): It is one of the simplest of the similarity measures which is calculated by subtracting pixels within a square neighborhood between the reference image I1 and the target image I2 followed by the aggregation of absolute differences within the square window, and optimization with the winner-take-all (WTA) strategy. If the left and right images exactly match, the resultant will be zero.

$$\sum_{(i,j)\in W} |I_1(i,j) - I_2(x+i, y+j)| \qquad (1)$$

SSD (Sum of Squared Differences): In this method the differences are squared and aggregated within a square window and later optimized by WTA strategy. This measure has a higher computational complexity compared to SAD algorithm as it involves numerous multiplication operations and hence takes longer to generate results.

$$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i, y+j)\right)^2 \qquad (2)$$

NCC (Normalized Cross Correlation): It is even more complex than both SAD and SSD algorithms as it involves numerous multiplication, division and square root operations. It generates the most refined output than SAD & SSD but takes more time for computation.

$$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i, y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i, y+j)}} \qquad (3)$$

SAD by Derivatives: We improved our original algorithm to account for derivatives along x-axis. A threshold value is assumed and differences in neighboring pixels greater than threshold will only be accepted for computation.

If I(i,j)-I(I,j-1)>threshold then follow SAD steps

Along with matching cost computation in order to obtain more accurate results we select windows for e.g 3X3, 5X5, 9X9. Depending on the window size selected the accuracy of the disparity map generated and the time taken to complete the computation is dependent. This windowing method deals with calculating correspondence for each and every pixel inside the window and then calculating correspondence for whole image.

Hence, it also includes aggregation step in which aggregation of results of every pixel is done in order to achieve better disparity results.

IV. EVALUATION METHODOLOGY

Quality measures are one most important standard to measure the accuracy of the algorithm by comparing the results with existing ground truths. To evaluate the performance of a stereo algorithm or the effects of varying some of its parameters, we need a quantitative way to estimate the quality of the computed correspondences. Two general approaches to this are to compute error statistics with respect to some ground truth data. The two methods considered in this paper are RMS error (Root Mean Square error) and BAD PIXEL Match.

1. RMS (root-mean-squared) error: (measured in disparity units) between the computed disparity map dC(x,y) and the ground truth map dT(x,y)[13], i.e. *R= sqrt [1/N∑(x,y) [dC(x.y)-dT(x,y)]]* (4)

where N is the total number of pixels.

2. Percentage of bad matching pixels: This quality metric gives percentage of mismatching pixels in the where λd (eval bad thresh) is a disparity error tolerance. For the experiments in this paper we use λd =1 .0, since this coincides with some previously published studies [15] and [16].

CUDA capable GPU: CUDA is NVIDIA's parallel computing architecture that enables huge increases in computing performance by harnessing the power of the

V. SIMULATION RESULTS

Stereo images are taken from online Middlebury Stereo Dataset along with their ground truths. Disparity maps for SAD, SSD, NCC and SAD by derivatives local stereo matching algorithms are computed for those Middlebury datasets [14] and our results (non-rectified) are compared with their ground truths available online to see how accurate are the results obtained. Our own data base results are also displayed .The actual distance

computed disparity map and the ground truth. $B = [\ 1/N\sum(x,y)\ [dC(x.y)-dT(x,y)\ ] > \lambda d\ ]$ (5)

GPU (graphics processing unit).CUDA capable GPU's allow to do Parallel programming, it allows to launch multiple parallel Threads which speed up the computation. To run a CUDA program one must need a machine running a CUDA capable GPU. The GPU model that is used is QUADRO K1100M with CUDA toolkit 7.5.

was measured and the Digital Camera with specifications-Brand Sony, Product Line Sony Cyber-shot, Model DSC-W220, Sensor Resolution 12.1 Megapixel Optical Sensor Size 1/2.3" was used to generate data base with 50 mm cameras apart.

Figures and Tables are shown below which show quantitative comparison of these methods among each other and online available datasets.

(A) Teddy Database (2003)

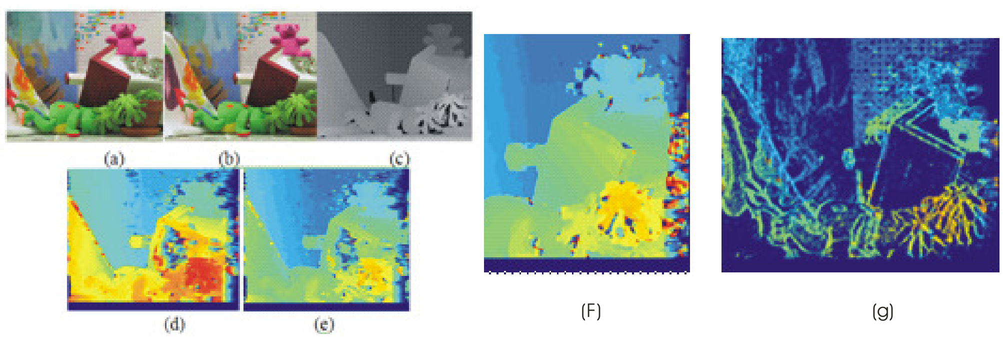

Fig. 1. TEDDY Dataset (a) and (b) input left and right images, (c) ground truth, (d) SAD disparity, (e) SSD disparity, (f) NCC disparity, (g) SAD by Derivatives disparity.

(B) CONES DATABASE(2003)

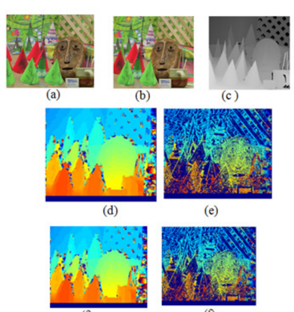

Fig. 2. CONES Dataset (a) and (b) input left and right images, (c) ground truth, (d) SAD disparity, (e) SSD disparity, (f) NCC disparity,(g) SAD by Derivatives disparity.

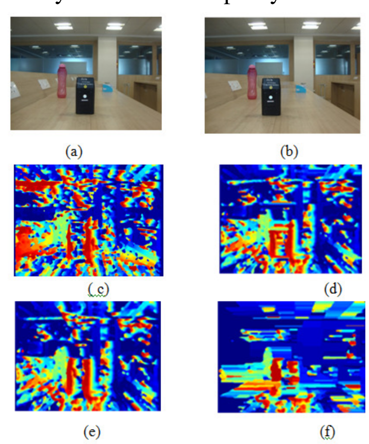

Fig. 3. OUR DATABASE (a) and (b) input left and right images, (c) SAD disparity (d) SSD disparity (e) NCC disparity (f) SAD by Derivatives disparity.

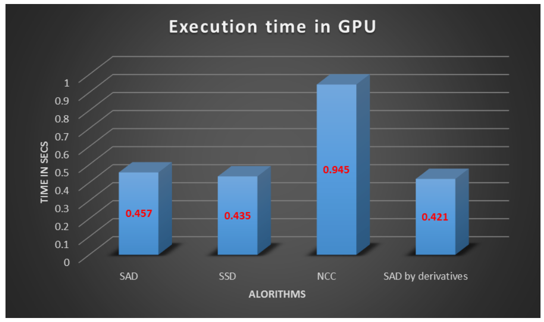

Fig. 4. Execution time of all algorithms in secs using multicore processors.

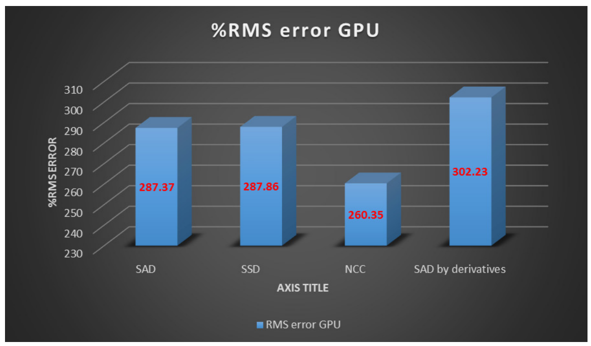

Fig. 5. Percentage RMS error of all algorithms using multicore processors.

Table 1: Comparison of local stereo matching algorithms using quality metric(BAD Pixel Match) for CONES dataset.

Method	Our algorithms bad pixel match values	Middlebury existing bad pixel match values
SAD	.041	0.11
SSD	.040	0.11
NCC	.038	0.09
SAD by derivatives	.047	--

VII. CONCLUSION

Stereo Matching Algorithms like SAD (sum of absolute difference), SSD (Sum of Squared Differences), NCC (Normalized Cross Correlation), SAD by Derivatives has been implemented to generate disparity Map. Time taken by each method to generate the disparity map is plotted in a bar graph shown as figure 4. Two quality metrics RMS error and BADPIXEL match are computed for each algorithm and the percentage RMS error is computed for the images and plotted as a bar graph figure 5.

Table I shows the comparison of our algorithms with existing Middlebury bad pixel match values. It gives 200% better performance than Middlebury dataset as computed. From all tables and figures we come to a conclusion that NCC takes more computation time BUT gives more accurate matching when compared with other matching techniques.

Non parametric methods work approximately same for normal scenes but work very fine for texture-less surfaces where parametric algorithm fail. It can also be concluded that SAD by Derivatives method takes very less time for computation but RMS Error and BAD pixel percentage value is close to other methods hence this algorithm can be used in situations where accuracy can be slightly compromised. Hence we come to a conclusion that there is a trade-off between computation time taken and accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank and show gratitude to Dr. D.Y. Patil Institute of Engineering, Research and Management, Akurdi for sharing their pearls of wisdom with us during the course of this research. We thank our colleagues who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

REFERENCES

[1] Benjamin, J. M. Jr. (1973). The Laser Cane. Bulletin of Prosthetics Research, pp. 443-450.

[2] M. J. Hannah. Computer Matching of Areas in Stereo Images. PhD thesis, Stanford University, 1974

[3] Schuon, S., Theobalt, Ch., Davis, J., & Thrun, S. (2008). High-quality scanning using time-offlight depth superresolution. In: IEEE CVPR Workshop on Time-Of-Flight Computer Vision 2008, pp. 1-7

[4]ouglas, T. S., Solomonidis, S. E., Sandham, W. A., and Spence, W. D. (2002). Ultrasound image matching using genetic algorithms. Medical and Biological Engineering and Computing, Vol.40, pp. 168-172.

[5] François, A. R. J. and Medioni, G. G. (2001). Interactive 3D model extraction from a single image. Image and Vision Computing, Vol.19, pp. 317-328.

[6] Nagai, T., Naruse, T., Ikehara, M., & Kurematsu, A. (2002). Hmm-based surface reconstruction from single images. In. Image Processing. 2002. Proceedings. 2002 International Conference on, Vol.2, pp. II-561 - II-564

[7] Ozden, K. E., Schindler, K., and van Gool, L. (2007). Simultaneous Segmentation and 3D Reconstruction of Monocular Image Sequences. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp. 1-8.

[8] Helmi, F. S. & Scherer, S. (2001). Adaptive Shape from Focus with an Error Estimation in Light Microscopy. 2nd Int'l Symposium on Image and Signal Processing and Analysis, pp. 188-193.

[9] Mingxiang, L., & Yunde, J. (2006, Dec). Trinocular cooperative stereo vision and occlusion detection. IEEE International Conference on Robotics and Biomimetics, 1129-1133

[10] Nalpantidis, L., Chrysostomou, D., & Gasteratos, A. (2009, December). Obtaining reliable depth maps for robotic applications with a quad-camera system. In International Conference on Intelligent Robotics and Applications (vol. 5928, p. 906-916). Singapore: Springer-Verlag

[11] T. Kanade. Development of a video-rate stereo machine. In ImageUnderstandingWorkshop, pages549–557, Monterey, CA, 1994. Morgan Kaufmann Publishers

[12] M. J. Hannah. Computer Matching of Areas in Stereo Images. PhD thesis, Stanford University, 1974

[13] T. W. Ryan, R. T. Gray, and B. R. Hunt. Prediction of correlation errors in stereo-pair images. Optical Engineering, 19(3):312–322, 1980.

[14] Williams, J. & Bennamoun, M. (1998). A Non-linear Filtering Approach to Image Matching. In: Proceedings of the 14th International Conference on Pattern Recognition, Vol.1, No.1, p. 3.

[15] R. Szeliski and R. Zabih. An experimental comparison of stereo algorithms. In International Workshop on Vision Algorithms, pages 1–19, Kerkyra, Greece, 1999. Springer

[16] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. IEEE TPAMI, 22(7):675–684, 2000.]