

A LOW-COST EMBEDDED STEREO VISION SYSTEM FOR REAL-TIME APPLICATIONS USING RASPBERRY PI

Dr. G.R. Gnana King & Ms. shilpa prakash ¹Associate Professor,ECE Sahrdaya College of Engineering & Technology, Kodakara kings.326@gmail.com ²M. Tech scholar Sahrdaya College of Engineering & Technology, Kodakara shilpaprakashottaly@gmail.com

Abstract— stereo vision is a widely applicable technology for extracting depth information. Depth is obtained in the form of disparity map from a pair of stereo images of the observed scene, which is captured using a camera pair. The technique of inferring depth information is called stereo matching and it is the key element in stereo vision. There are different stereo matching algorithms which are broadly classified into global algorithms and local algorithms. In this paper stereo vision system based on semi global block matching algorithm and its low cost implementation is proposed. It results in a standalone real_time stereo vision system for low cost and high performance embedded applications. Moreover, the efficiency of the algorithm is further improved by using additional pre and post processing steps. The proposed stereo vision system can be used in autonomous navigation, robot vision and object tracking etc.

Keywords — stereo vision, stereo matching, triangulation, disparity map, raspberry pi, openCV.

I. INTRODUCTION

Stereo vision or computer vision is the technology that is used to extract the depth information from a pair of stereo images. It is similar to human binocular vision where the depth information is inferred in our brain after processing the image pair captured by our eyes. similarly the depth information about the scene is obtained by capturing it using two similar cameras placed on the same horizontal surface at a fixed distance. The technique used for getting depth information is called stereo matching. It has wider application in aerial survey to autonomous navigation ,robot vision etc.

Stereo vision is an active research area, as a result number of stereo vision algorithms have been developed. They are broadly classified into local and global algorithms.local algorithm uses The local features such as intensity level, color values etc for determining the disparity map. The advantage of local algorithm over global is that it require minimum memory and its computational complexity is very less. In contrast to this global algorithm use global assumptions about energy or smoothness to determine disparity map. It require more memory and its computational complexity is very high. The basic principle of stereo matching is called triangulation.



Fig 1. triangulation principle

In the triangulation principle two cameras or the observing points will be placed on the same horizontal line as shown in fig 1.1. These points are separated by a fixed distance b.where Z_l and Z_r are left and right optical axis, where as X_r and X_l are the focus planes respectively. Now let M be the image point, then the depth is obtained from the equation (1), where the disparity(ml+mr) and depth are inversely proportional.

depth=fb/ml+mr

(1)

In this paper we are using semi global algorithm(SGM). It is one of the top performing algorithm.It successfully combines the concepts of global and local stereo methods for accurate, pixel-wise matching at low runtime. In contrast with the existing algorithms SGM does the optimization across the image.Along with this pre and post processing steps are done in order to improve the efficiency of the algorithm.

The whole system is implemented using raspberry pi in python using opencV platform. Real time Stereo images are captured using pair of web cameras.

II REVIEW OF STATUS OF RESEARCH AND DEVELOPMENT

Stereo images captured using camera pair gives the depth information of the scene in the form of disparity map.The technique used for obtain the depth information is called stereo matching.It is formulated as the problem to



find correspondence pair. Pixel in the left image is compared with corresponding epipolar line pixels in the right image.Each and every pixel along the epipolar line is examined and the best match is identified .Searching for correspondence pixel along the epipolar line rather than searching in the 2d region reduce the difficulty in finding the correspondence pixel and complexity of the algorithm .Stereo vision algorithms are classified as local algorithms and global algorithms based on feature they used to generate the disparity map.Stereo matching algorithms generally have 4 major steps 1.Cost estimation 2.Cost aggregation 3.Disparity estimation and 4.Disparity refinement[15].



fig 2.general stereo vision steps

In local algorithms all this four steps will be there but in global method cost aggregation method is optional. Cost matching measures the similarity between two pixels. It is in this stage the values of two pixels correspond to the same point in a scene is determined. So we can say that this is the stage where the correspondence of image points in left and right image is determined[16]. The intensity difference between the matching pixel gives the disparity value.Cost measurement can be done using area based or window based technique. These technique produce more accurate results as they consider entire set of pixels associated with image regions.Most commonly used algorithms for window based technique are the sum of absolute differences (SAD), the sum of squared differences (SSD), normalized cross correlation (NCC), rank transforms (rt), and census transforms (CT) [17].Cost aggregation is an optional step for global algorithms but it determines the performance of local algorithms. The use of cost aggregation is to minimize the matching uncertainty .In local method matching cost aggregation is done by summing them over the support region. Usually the support regions will be square window with the pixel of interest at the center[18].Disparity optimization or computation is the step in which the final disparity is selected .In case of local algorithm the final disparity is determined by winner takes all method(wta). It is defined by

$$d_p = \arg_d \epsilon D \min C'(p,d).$$
(2)

where d_p represent the disparity value associated with minimum aggregate cost of each pixel.aggregation cost obtained after the matching cost calculation is represented by C'(p,d).The set of all discrete disparity is represented by D.this method is used in papers[19],[20] and [21].According to them the disparity map obtained from this contain errors in the form of unmatched pixels or occluded region .This errors are removed in the disparity refinement steps by using filters.in contrast ,global method which determines the disparity map by the assumptions about energy minimization. The mostly used assumption is that the scene is locally smooth except for object boundaries, and thus neighboring pixels should have very similar disparities.in such case the main objective will be finding an optimal energy disparity assignment function d = d(x,y) that minimizes

$$E(d) = E_{data}(d) + \beta E_{smooth}(d)$$
(3)

In equation(2) $E_{data}(d)$ represents the matching costs at the coordinates(x,y), $E_{smooth}(d)$ the Smoothness energy encourages neighboring pixels to have similar disparities based on the previous stated assumptions and β is a weighting factor. The disparity refinement step reduce the noise and improves the disparity map.Filters are used for this.In local method typical disparity refinement steps include gaussian convolution and median filter[22],[23]. Broadly the stereo matching techniques are classified into global and local algorithms. Where local algorithms determines disparity that is associated with a minimum cost function at each pixel by performing block matching and winner takes all optimization.Local methods have less computational complexity and memory usage. The local methods also makes use of either fixed windows or multiple windows with different size. The demerit of local method is that the accuracy is reduced as compared with the global algorithms.In global algorithm ,which are formulated as energy minimization problems ,the computational complexity and memory usage is high. The global algorithms mainly use technique such as dynamic programming, graph cuts and belief propagation etc for solving the energy minimization problem

In order to incorporate the merits of the both global and local algorithm proposed system make use of semi global matching algorithm. It produce more accurate results with better runtime and less memory usage. Along with the use of this semi global algorithm in order to improve the image processing pre and post processing steps are also included. The efficiency of the algorithm is tested by implementing the system using raspberry pi on opencv platform using a pair of camera sensors that captured the input left and right stereo images in real time.

III PROPOSED SYSTEM

Stereo vision provides the depth information of the observed scene in the form of disparity. It is formulated as the problem of finding correspondence in left and right images.Proposed system use semi global matching algorithm along with some pre and post processing steps .The real time implementation is done in raspberry pi and a pair of camera using opency python. Section A We discuss the pre and post processing,section B Sgm and section C ,Implementation.



Available at https://edupediapublications.org/journals



fig 3.proposed system

A. Pre and post processing

The proposed stereo vision system is based on SGM Algorithm.In order to improve the accuracy of the disparity map, pre and post processing steps are done. Since system is implementing in real time the quality of the capturing images depends on the environment conditions such as light intensity ,because of this captured images are pre processed.Preprocessing steps are done for removing noise and for contrast enhancement .Noise in the captured images are removed by using median filter .After the preprocessing and processing ,postprocessing steps are done .Postprocessing steps includes outlier suppression and peak removal.fig 3 shows the difference in disparity map with and without pre and post processing steps.



fig 3.i)without preprocessing ii)with preprocessing

B. Semi Global Matching Algorithm

Semi global matching is one of the top performing stereo matching algorithm .It combines the concept of both local and global algorithm.It has the merits of both classification.It is prefered over the other existing algorithm because it provides a good trade off between runtime and accuracy ,mainly at object borders and time.In every algorithms the major steps are matching cost calculation ,aggregation and disparity determination.In SGM pixelwise cost matching and aggregation by consider 2d smoothness constraint that combines many 1d constraint is done.

1) Pixelwise Cost Matching :

In SGM cost is calculated pixelwise.Consider the pixel V of the left image.The matching cost is calculated from the intensity I_{bv} of the pixel V and its correspondence I_{mw} at w.The pixel W can be expressed as a function $w = e_{bm}(V,d)$ since w is along the epipolar line e of the left image pixel V with a line parameter d ,where d represent disparity.In case of rectified image

$$e_{bw}(V,d) = (V_x - d, V_y).$$
(3)

In pixel cost matching the method used is proposed by birchifield and tomari[1] in which cost calculation of the pixel is the absolute difference of intensities at V and w=ebm(v,d) which is in the range of half a pixel in each direction along the epipolar line. This method matches the individual pixels in one image with its correspondence pixel in the pair image. It gives disparity map that includes the sharp changes in disparity while introducing a few false discontinuities.

In this method each match in match sequence is an ordered pair (v,w) for which $I_{l}(v)$ and $I_{r}(y)$ are the intensity of the intensities of same image points in the pair images. The unmatched pixels are occluded, where occlusion is for those image points that are not present in one image of the pair but not in the other image. While calculating the disparity d of a pixel V in the left image that Match some pixel w in the right image is defined as v-y. The depth discontinuity pixels are considered for now to be those pixels that border a change of at one disparity level and lies on the far object.

The match sequence is the way in which correspondence is encoded.The cost of this is defined as a constant penalty for each occlusion.

$$\gamma(z) = N_{occ} K_{occ} - N_m K_r + \sum_{i=1}^n d(V_i, W_i)$$
(4)

Where z represent the each match segment and γ (z) is the associated cost that gives the measure of how much dissimilarities are there between m and the true correspondence, where m is the match sequence.

2) Aggregation of cost :

Calculation the pixelwise cost is generally ambiguous and wrong matches can easily have a lower cost than correct ones, due to noise, etc.In order to avoid this, an additional constraint is added that supports smoothness by penalizing changes of neighboring disparities [2].

In equation (5) E(D) represent energy that defines the pixel cost and smoothness constraints. It depends on the disparity image D.

$$\varepsilon(D) = \sum_{p} c(p, Dp) + \sum_{w \in Nv} V T \left[\left| Dv - Dw \right| = 1 \right] + \sum_{w \in Nv} V 2T \left[\left| Dv - Dw \right| > 1 \right]$$
(5)

The first term is the sum of all pixel matching costs for the disparities of the disparity image D. The second term and the third terms adds a constant penalties.penalty V 1 Is added for all pixels w in the neighborhood N_{-v} of V, for which the disparity changes a little bit (i.E. 1 Pixel).in order for larger disparity changes the third term adds a larger constant penalty V2 .by assigning a lower penalty for small changes permits an adaptation to slanted or curved surfaces whereas the addition of constant penalty for all larger changes Independent of their size preserves discontinuities . Discontinuities are often visible as intensity changes. This is exploited by adapting V2 to the intensity gradient, i.e. V 2 = $|i - ibv \ bw |$ however, it has always to be ensured that V2 \geq pv1.

in order to find the disparity image D that minimizes the energy E(D) is the formulation of the stereo



matching problem . but for many discontinuity preserving energies such a global minimization (2d) is np-complete. In contrast, the minimization along individual image rows (1d) can be performed efficiently in polynomial time using dynamic programming . However, dynamic programming solutions easily suffer from streaking , due to the difficulty of relating the 1d optimizations of individual image rows to each other in a 2d image. The problem is, that very strong constraints in one direction ie. along the image rows , are combined with none or much weaker constraints in the other direction ie. along image columns.

a new idea is made for aggregating matching costs in 1d from all directions equally. for a pixel V and disparity d smoothed cost s(v, d) is calculated by summing minimum cost paths which is along all the id that end in pixel p at disparity d in order for calculation the cost along the path is considered not the path itself.consider the equations (6),(7),(8), where t r be a path that is traversed in the direction r. The cost t r (v,d) of the pixel p at disparity d is defined recursively as,

(7)

$$T r (v,d) = c(v, d) + min(T r (V-r, d), (6))$$

$$T r (v-r d - 1) + V1, T r (V-r, d + 1) + V1$$

$$Mini T r (V-r, i) + V2)$$
(8)

Mini $T_{r}(V-r, i) + V2$) (8) As a penalty for discontinuities low cost of the previous pixel p - r of the path is added to the remainder of the equation which implements the behavior of equation along an arbitrary 1d path. this approach is more likely a scanline optimization than traditional dynamic programming solutions as the cost does not enforce the visibility or ordering constraint, these concepts for paths that are not identical to epipolar line cannot be realized. due to this values of t increase s along the path permanently ,which may lead to very large values. However, by subtracting the minimum path cost of the previous pixel from the whole term THE equation can THUS BE modified.

$$T_{r}(v, d) = c(v, d) + \min(T_{r}(V-r, d), (9) T_{r}(V-r, d+1) + V1 (10) Min tr(V-r, i) + V2) - \min tr(v-r, k)$$
(11)

The subtracted value is constant for all disparities of a pixel p so that the modification does not change the actual path through disparity space and the position of the minimum remained as such. However, $\leq c \max + V2$ is given as the upper limit and the costs tr are summed over paths in all directions r.in order to provide a good coverage of the 2d image number of paths should be 16 or at least it must be 8 ,if coverage is compromised.

$$S(v, d) = \sum_{r=1}^{r} L_{r} T_{r}^{r} (v, d)$$
(12)

for s the upper limit is easily determined as s $\leq 16(c \max + V2)$.

3) Disparity estimation :

In order for the disparity estimation the method mentioned in [2] is used. Each pixel v is selected with

disparity d that corresponds to minimum cost ie Min d s(v, d). This selected pixels are used for determining db that is correspondence to ib,the base image. For the estimation of sub-pixels the sum of squared differences is used for finding correlation. It is used because of the simplicity of calculation. same costs of pixels that lies along the epipolar line of w of the image bm is used to determining disparity image dm that correspond to match image im .i.e Min d s(e mb (w, d), d).

In this cost aggregation step, for getting a better result the ib and im are not treated symmetrically. Whenever dm is calculated from scratch a median filter with a window size 3x3 is used for filtering outliers from disparity images of base and match images. Determination of occlusion and false matches are done by consistency check which is made possible due to the calculation of disparity images dm and db.For the purpose of finding valid disparity ,the disparity images of base and match images are compared.If they match ,they are valid.

$$D_{nv} = d bv if |d bv - d mw| \le 1, q = e bm (v, d bv)$$

otherwise.
(13)

One to one mappings is used in order to do consistency check that enforces the uniqueness constraint.

C. Implementation

The SGM stereo matching algorithm is implemented on raspberry pi using opencv platform in python language.For real time processing the input images are captured using a pair of camera having same specification.They are placed along same horizontal line with a fixed distance between them and the output ,which is the depth in the form of disparity map is showed in a display unit.

VI .EXPERIMENT AND RESULTS

In order to evaluate the efficiency of proposed algorithm, at first the database images (450x375 p) were tested and a satisfactory results were obtained .After that the real time implementation of the system was done.In order for real time processing ,the input images ,both left and right were captured using a pair of camera with same specification and placed along a horizontal line with a fixed separation.

Images that are available in the middlebury dataset were used first for the testing. A set of stereo image pairs with ground truth were selected for evaluation purpose (first row of figure 2). All the images were processed with a disparity range of 32 pixel.

As shown in fig 2 the right and left images were processed using the proposed algorithm .Then the obtained results (iii)fig 2.) Were compared with the ground truth images.(Ii)fig2).

For improving the results ,the images captured using sensor



pair is preprocessed. It is evident that the proposed algorithm with and without preprocessing produced different output efficiency which remark the effect of preprocessing steps. The preprocessing steps including image enhancement and noise removal steps were done. The results obtained with and without preprocessing are shown in the fig 3.



Fig.2 i)the left images of teddy,shapes,cones ii)ground truth values iii)result of proposed algorithm.

After getting the disparity map their color maps were obtained in order to get more visual information about depth.

matching disparity map and SGM is shown below.



fig 4)i)block matching ii)proposed

The real time results are shown below. Input images that are sensed by camera pair showed different quality as they depend on the intensity of light in the observed scene and also the resolution of the sensors used .A pair webcam from logic tech ,3 mega pixel usb camera with frame rate up to 30fps and having video resolution:1280x720 pixel hd, were used as sensors.

Quality and speed of the proposed algorithm was tested with the results of the related works .The processing speed of the proposed sgm was measured in frames per second and in million disparity estimations per second (mde/s), it is a metric that also takes into account the number of pixels and the disparity range.In terms of speed in mde/s comparing with the related works our proposed give comparatively fast and and accurate disparity map.Which also implemented at a low cost .



Fig 3. i) left image and right image ii)disparity map gray scale iii)color map

The most commonly used algorithm for disparity estimation is called block matching comparison of block



Available at https://edupediapublications.org/journals



iv

fig 5.i)real time inputs ,ii)threshold,iii)gray scale disparity iv)color map.

In terms of speed in mde/s comparing with the related works our proposed give comparatively fast and and accurate disparity map. Which also implemented at a low cost .The image processed was having resolution of 1280x720. Metric mde/s is calculated usingmde/s =m.D.N.Fps. Where d is the disparity, fps stands for frames per second and n is the number of pixels.

VI CONCLUSION

Stereo vision systems are now used in many fields such as autonomous navigation, robotics etc and its application is extended to other fields too. In this paper a stereo vision system for accurate disparity map generation is proposed. The system is based on semi global block matching with improved pre and post processing operations. The system is reliable and low in cost. The system is implemented using raspberry pi using opency platform. The advantage of this system over the existing ones is that it gives more accurate results at low runtime and low implementation cost.Future research will be applied to additional indoor and outdoor scenes and also for processing video.

quanty and speed medsurement									
work	average error rates				image	D	spee d(fps	MD	platform
	NONOC C	AL L	DIC	OVERAL L	size	2)	E/s	P
Baha[13]	7.16	11.5	n.a	n.a	450x375	50	46	388	FPGA
Zaha[14]	4.141	7.41	12.8	8.20	1024x76 8	64	60	3019	FPGA
Ambros h[12]	5.82	9.37	22.2	12.5	750x400	60	60	1080	FPGA
Ma[15]	n.a	n.a	n.a	n.a	450x375	59	45	488	CPU/GP U
perri[16]	n.a	n.a	n.a	n.a	512x512	255	25.6	1711	FPGA
propose d	2.6	n.a	10.3	n.a	1280x72 0	64	30	3200	CPU/Pi

TABLE 1 quality and speed measurement

REFERENCES

[1]S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. In Proceedings of the Sixth IEEE International Conference on Computer Vision, pages 1073–1080, Mumbai, India, January 1998.

[2]Hirschmuller, H. (2005, June). Accurate and efficient stereo processing by semi global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, pp. 807-814). IEEE.

[3]Hamzah, R. A., & Ibrahim, H. (2015). Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, 2016.

[4]M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 993– 1008, 2003.

[5]N. Lazaros, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," International Journal of Optomechatronics, vol. 2, no. 4, pp. 435–462, 2008.

[6]D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence



algorithms," International Journal of Computer Vision, vol. 47, no. 1–3, pp. 7–42, 2002.

[7]C. Cigla and A. A. Alatan, "Information permeability for stereo matching," Signal Processing: Image Communication, vol. 28, no. 9, pp. 1072–1088, 2013.

[8]K. Zhang, J. Lu, Q. Yang, G. Lafruit, R. Lauwereins, and L. Van Gool, "Real-time and accurate stereo: a scalable approach with bitwise fast voting on CUDA," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, pp. 867–878, 2011.

[9]Z. Lee, J. Juang, and T. Q. Nguyen, "Local disparity estimation with three-moded cross census and advanced support weight," IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 1855–1864, 2013.

[10]Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 3, pp. 492–504, 2009.

[11]Y. S. Heo, K. M. Lee, and S. U. Lee, "Joint depth map and color consistency estimation for stereo images with different illuminations and cameras," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1094–1106, 2013.

[12]K. Ambrosch and W. Kubinger, "Accurate hardwarebased stereo vision," Comput. Vis. Image Understanding, vol. 114, no. 11,pp. 1303–1316, Nov. 2010.

[13] N. Baha and S. Larabi, "Accurate real-time neural disparity MAP estimation with FPGA," Pattern Recog., vol. 45, no. 3, pp. 1195–1204, Mar. 2012.

[14] L. Zhang, et al., "Real-time high-definition stereo matching on FPGA," in Proc. 19th ACM/SIGDA Int. Symp. Field Programmable Gate Arrays, Monterey, CA, USA, 2011, pp. 55–64.

[15]S. Perri, D. Colonna, P. Zicari, and P. Corsonello, "SAD-based stereo matching circuit for FPGAs," in Proc. 13th IEEE Int. Conf. Electron., Circuits Syst., 10-13 Dec. 2006, pp. 846–849.

[16]Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in Proc. Int. Conf. Comput. Vis., 2013, pp. 49–56. [17]S. Mattoccia, S. Giardino, and A. Gambini, "Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering," in Proceedings of the Asian Conference on Computer Vision (ACCV '10), vol. 38, pp. 371–380, Xi'an, China, September 2010.

[18]Á. Arranz, Á. Sánchez, and M. Alvar, "Multiresolution energy minimisation framework for stereo matching," IET Computer Vision, vol. 6, no. 5, pp. 425–434, 2012.

[19]L. Xu, O. C. Au, W. Sun et al., "Stereo matching by adaptive weighting selection based cost aggregation," in Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '13), pp. 1420–1423, Beijing, China, May 2013

[20]J. Salmen, M. Schlipsing, J. Edelbrunner, S. Hegemann, and S. Luke, "Real-time stereo vision: making more out of dynamic programming," in Computer Analysis of Images and Patterns, pp. 1096–1103, Springer, Berlin, Germany, 2009.

[21] K. R. Vijayanagar, M. Loghman, and J. Kim, "Realtime refinement of kinect depth maps using multi-resolution anisotropic diffusion," Mobile Networks and Applications, vol. 19, no. 3, pp. 414–425, 2014.

[22]M. Michael, J. Salmen, J. Stallkamp, and M. Schlipsing, "Real-time stereo vision: optimizing semi-global matching," in Proceedings of the IEEE Intelligent Vehicles Symposium (IV '13), pp. 1197–1202, Gold Coast, Australia, June 2013.

[23]J. M. Pérez and P. Sánchez, "Real-time stereo matching using memory-efficient Belief Propagation for highdefinition 3D telepresence systems," Pattern Recognition Letters, vol. 32, no. 16, pp. 2250–2253, 2011.