Disparity refinement through grouping areas and support weighted windows

Gabriel da Silva Vieira[1*§], Fabrizzio Alphonsus A.M.N. Soares[2*], Gustavo T. Laureano[3*],
Rafael T. Parreira[4*], Júlio C. Ferreira[5§], Ronaldo M. Costa[6*], Cristiane B. R. Ferreira.[7*]
[*]Federal University of Goiás, *Pixelab Laboratory*. Goiânia - GO, Brazil.
{fabrizzio[2], gustavo[3], ronaldocosta[6], cristiane[7]}@inf.ufg.br, rafaeltp3[4]@gmail.com
[§]Federal Institute Goiano, *Computer Vision Laboratory*. Urutaí - GO, Brazil.
{gabriel.vieira[1], julio.ferreira[5]}@ifgoiano.edu.br

***Abstract*—In this work, we propose a simple but an effective technique to adjust a disparity map in a more appropriate configuration. This proposal consists of three main steps: segmentation process, statistical analysis and by using adaptive weighted windows. Furthermore, we investigate if a disparity map, yielded by a robust stereo method, can be improved by the proposed methodology. Thus, we implement some stereo vision methods to compare. The experimental results show that the proposed method is efficient and it can make some enhancements in disparity maps, as reducing the disparity error measure.**

Keywords: stereo vision, image segmentation, adaptive support window, disparity map, map adjustment, disparity methods.

I. INTRODUCTION

In stereo vision systems, most algorithms are organized in four steps: (1) *matching cost*, (2) *cost aggregation*, (3) *disparity selection* and (4) *disparity refinement*. This pipeline was pointed out by [1] and a lot of work has been done based on it.

Our study focuses on step 4 of the pipeline, which is the *disparity refinement*. In this work, we propose a simple but an effective technique to adjust a disparity map in a more appropriate configuration. It is based on an assumption that disparity map regions carries helpful information. As a set of disparity are allocated in a region, the disparity with more frequency, i.e, the disparity that mostly appears in a region, indicates the correct one. Hence, other disparities in that region can be discarded. The next step is to find the disparity of these points. To do that a weighted function is performed for this task.

This proposal consists of three main steps. Firstly, a segmentation process is applied to correlate similar points and, in this way, to distinguish non-equivalent points, each one in its own segment. Secondly, for each segment, a statistical analysis is performed and the values returned by a mode function are used to remove wrong disparities. Finally, to fill the disparity map in, a plausibility given by a adaptive weighted window is used to replace unknown disparities.

As this strategy uses the weighted window technique, we investigate if it can improve results provided by stereo vision methods that also use this technique. Thus, we implemented some stereo methods and we took their outputs as a raw disparity map to be improved by the proposed methodology.

By performing an evaluation, we can observe that this method can improve disparity maps substantially. Even maps with a good accuracy can be enhanced by this method. In addition, we pointed out that although this proposal is presented in the context of local approaches, it can also be applied in global strategies in the same way.

The remainder of the paper is organized as follows. After briefly reviewing closely related work in Section II, we show a overview of the technique proposed and we describe our algorithm for adjusting a disparity map in Section III. The experimental results and analysis are given in Section IV and Section V concludes the paper.

II. RELATED WORK

In local stereo methods, an aggregating window is used to calculate the similarity among points by considering a neighborhood region. They consider the entire set of pixels associated with image regions that may be square or rectangular and may be fixed or adaptive in size [2].

This surrounding pixels area is commonly referred to as support or aggregating window and it is applied to get better results mainly in uniform areas. When each point of the support has a value that shows its strength, we are talking about support weighted functions.

By considering that points in a window have different influencies, a support weighted function is performed to model this behavior. Yoon and Kweon [3] have started this analysis and they showed that photometric and geometric constraints can be applied to identify each weight.

In compare with other local methodologies, a disparity map can be prepared with high accuracy through support weighted windows. This is because a color and space proximity can emphasize similar regions, consequently improving the matching cost step.

Laureano and Paiva [4] proposed a method based on support weighted window. In their approach a multi-resolution analysis is performed, thus the matching cost procedure is applied in each pyrimid level and its result is propagated for the other levels. This method deals in particular with textureless regions.

Rhemann et al. [5] considered a guided filter to define the weights of the support. Their method shares the edge-preserving property with the joint bilateral filter and because of that it achieves great results.

Gerrits and Bekaert [6] focused in outlier rejection during the aggregation step. Their method uses a segmentation process and based on that it suggests a weight for each point that lie in the same segment of the window's center pixel and another weight for points outside of the segment of the central pixel.

Hosni et al. [7] investigated the support weighted window by performing some analyses and by comparing different

978-1-5386-2410-4/18/$31.00 ©2018 IEEE

approaches. One of their conclusion is that spatial constraint can be omitted when the weights are being calculated and this approach make very little difference on the quality of results.

In this brief section, some support weighted methods were pointed out. We use them to perform an evaluation that analysis the performance of the proposed method. Thus, for do that, we use our own implementation of these methods.

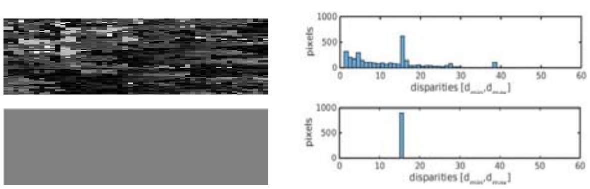
III. Proposed Approach

We start by analysing a raw disparity map. Fig. 1 shows a map that is very noisy in some parts of it. It was made by a simple cost aggregating (CA) methodology that can be called as *fixed window* (FW) method. It is the simplest CA strategy that uses an aggregating window and it is at the foundation of stereo vision systems. Besides, this map was also yielded by using a simple cost function that is the *sum of absolute differences* (SAD).

(a) (b)

Fig. 1: Disparity maps: (a) a raw map and (b) a ground truth map.

Fig. 1a illustrates a region that has a group of wrong disparities. Similar pixels coexist in this area and because of that, FW method fails in a lot of points. However, when we analyse these disparities we can see that most of the values are pointing to a correct one. Fig. 2 shows an histogram plot which confirms our analyse by comparing this map region with the same region in the reference map (*ground truth*), Fig. 1b.

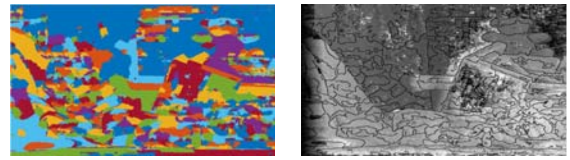


(a) (b)

Fig. 2: Map analysis: (a) disparity map regions and (b) related histograms.

In this way, if in a certain region a disparity method hits more than fails, we can use it. Unfortunately, it is something that we don't know because the correct disparity is still unknown. But if we believe in it, we can propagate this supposed correct value even knowing that this is not true all the time. Our methodology starts with this belief.

To identify a region, a segmentation technique may be used. In stereo vision systems, mean shift algorithm [8] is widely employed. It was used to obtain great results in [6], [9] and [10]. We use it to apply a segmentation in the reference image. When we obtain these segments we use them to localize regions in the disparity map. Fig. 3 shows a segmented image and its corresponding disparity map labeled based on these segments.



(a) (b)

Fig. 3: Image segmentation: (a) reference image and (b) disparity map.

After that, the method calculates the most common value for each segment. It is a simple equation that is show in Eq. 1. For each segment S with the identifier i into the disparity map D, $(S_i \subset D)$, it calculates the mode of all n segments and the results are stored in m.

$$m_i = mode\ common\ value\ in\ S_{i=1}^{n} \tag{1}$$

Moreover, each point of the disparity map that belongs a certain segment is evaluated, accordingly with the previous mode contability. In Eq. 2, a disparity value in D with the coordinates (x, y) is tested. In case of this value is in a range test, the mode value m is assigned for this point. Otherwise, it is assigned with 0 that represents a unknown disparity. In this equation, t is a threshould defined by a user that is used to approximate disparity values to the segment's mode. Besides, it considers that each disparity value is in a segment S with the identifier i.

$$D(x,y) = \begin{cases} m_i & \text{if } D(x,y)_{\in\{S_i\}} \in [m_i - t, m_i + t], \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

When applying the above equations, a disparity map is returned. At this time, disparities that are far away from their segment mode value are considered as unknown. The next step consists of filling these holes in, so a weighted filter is prepared to evaluate the plausibility of each possible disparity.

Yoon and Kweon [3] introduced a support weighted window to be applied in the stereo matching problem. Their methodology considers the color similarity between points and their space distance. A window is defined and a point located in the middle of this window is the principal point. The surrounding neighbors are compared with the principal point by calculating their difference of colors and their geometric distance. This strategy was used in [11], [12], [4] among others and investigated in [7].

The color proximity constraint between a principal point p and its neighbor point n within a support is given by:

$$f_c(\Delta c_{pn}) = e^{-\frac{\Delta c_{pn}}{\gamma c}} \tag{3}$$

The color distance Δc_{pn} represents the Euclidean distance between the colors of p and n in an image I as

$$\Delta c_{pn} = \sqrt{\sum_{j \in r,g,b} (I_j(p) - I_j(n))^2} \quad (4)$$

In the same way, spatial proximity constraint is evaluated accordingly to:

$$f_s(\Delta s_{pn}) = e^{-\frac{\Delta s_{pn}}{\gamma s}} \quad (5)$$

the spatial distance Δs_{pn} represents the Euclidean distance between the coordinates (x, y) of p and n as

$$\Delta s_{pn} = \sqrt{(p_x - n_x)^2 + (p_y - n_y)^2} \quad (6)$$

γc and γs refer to a constant of color similarity and a constant to adjust the spatial distance term, respectively. $f_c(\Delta c_{pn})$ and $f_s(\Delta s_{pn})$ represent the strength of grouping by color similarity and by proximity.

Color and spatial constraints are combined and the final support weighted window is given by

$$W(p, n) = e^{-(\frac{\Delta c_{pn}}{\gamma c} + \frac{\Delta s_{pn}}{\gamma s})} \quad (7)$$

In our method, we use the support weighted window with an adaptation. It is only applied in unknown disparities so a principal point in a window is a point o disparity that we want to discovery. Each neighboring pixel that has a disparity value is evaluated according to the previous equations. Thus, the weights of each pixel that are in the same disparity are accumulated. Fig. 4 helps in the explanation.

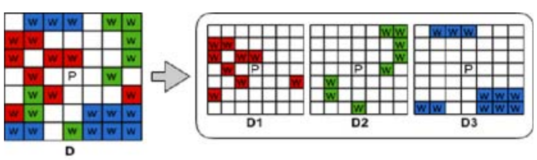


Fig. 4: Points in a disparity map D have their own weight w. Each segment is painted didactically (in red, green and blue). Weights in $D1$, $D2$ and $D3$ are summed separately. The best value is used to set the disparity in the principal point p.

Based on the color reference image, the photometric and geometric constraints are calculated and each point of the window has a weight w. Besides the weights, we know some disparities. In Fig. 4, each color represents a known disparity, except for a white color point that represents an unknown disparity and because of that these points don't have a weight w. Thus, the computed weights that are in the same disparity are summed up as in

$$\Omega_{d \in \{d_{min}, d_{max}\}} = \sum_{j=1}^{n} w_{ij} \quad (8)$$

where d_{min} and d_{max} are the range from minimum to maximum disparity and Ω is the accumulated sums. Hence, a disparity optimization is performed to select the best disparity. It is given by

$$D_{(x,y)} = \text{argmax}(\Omega) \quad (9)$$

where (x, y) are the coordinates of the unknown disparity in the disparity map D. Based on the best value from Ω, its disparity value is assigned to D.

Our method is inspired by considering a raw disparity map which has noisy parts. However, in this study, we investigate if a better disparity map, yielded by a robust stereo method, can be improved by the proposed methodology.

IV. Experimental Results

In this section, the results and the organization of the experiment are presented. Four image pairs were selected from the Middlebury dataset [13]. Each pair has its own ground truth that was used to evaluate the results. The methodology followed the specifications of [1].

In Table I, parameters ALL, NOCC and DISC are defined according to the Middlebury Stereo Evaluation - version 2 [13]. Although this version is no longer active, it is still being used, as in [14]. Some evaluation masks are provided and they are used to remove pixels that are not considered in the statistics. ALL is the error computed on the whole image, NOCC is the error computed on the whole image excluding the occluded regions and DISC is the error computed within the discontinuity regions [11].

In the test cases, raw disparity maps from Bilateral support weights (BL and BLNoSpatial), Multi-resolution and Perceptual Grouping (MRPG), Guided filter support weights (GF) and Segmentation-based (SB), are used as input for the proposed methodology, referred to as *segment consistency-check* (SCC).

We used a 39×39 support window to build raw disparity maps and the SAD cost as a measure of stereo matching. The color and spatial terms were set as $\gamma c = 23$ and $\gamma s = 14$. For the threshould in Eq. 2, it was set as $t = 1$.

Table I shows the accuracy of the proposed method. The SCC method decreased the percentage of bad pixels in the three considered parameters for Tsukuba image pair. In this test, for all stereo vision methods the error was reduced.

However, for Venus image pair the error increased in NOCC parameter and sometimes in the ALL parameter. It also happened in Teddy and Cones image pairs. Besides the characteristics of each scene, this is probably due to the segmentation process. We use only Tsukuba image to tune the parameters that define the segments. After that, we use them to segment the other image pairs. In spite of this, the SCC method was efficient to preserve discontinuity regions as shown in the DISC parameter.

A qualitative analysis shows the proposed methodology compared with the raw disparity maps. Because of the space limit, we show only the Tsukuba and Venus results. Fig. 5 shows the disparity maps to each image pair and the disparity map with SCC method. The first column from Fig. 5 corresponds to the raw disparity maps. In the second column, there are maps after applying Eq. 2. The third column shows the final result and the last column displays the ground truth map.

TABLE I: Accuracy evaluation.

Method	Tsukuba			Venus			Teddy			Cones		
	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC
BL [3]	4.68	5.96	19.58	6.05	7.49	32.61	15.78	24.29	32.59	8.65	18.25	20.49
BL + SCC	3.35	4.13	15.44	5.77	6.47	19.42	16.30	23.71	28.54	9.04	18.08	20.03
BLNoSpatial [7]	5.27	6.26	21.90	6.98	8.36	34.64	17.66	25.82	35.83	10.25	19.34	22.38
BLNoSpatial + SCC	4.04	4.80	17.97	7.70	8.60	26.53	17.52	24.73	31.28	10.80	19.52	22.44
MRPG [4]	3.30	5.12	13.50	0.97	2.48	8.99	10.58	19.66	22.15	5.90	16.18	14.12
MRPG + SCC	1.96	2.42	9.98	5.29	5.65	10.86	11.24	17.32	19.74	6.85	14.74	13.90
GF [5]	8.51	10.15	24.80	11.14	12.57	39.81	21.49	29.46	38.75	14.62	23.99	28.75
GF + SCC	4.00	5.06	15.98	6.01	6.70	21.98	16.91	24.10	29.41	10.88	20.23	23.37
SB [6]	3.79	4.33	15.31	5.46	6.62	28.24	15.30	22.18	29.76	12.57	21.03	22.87
SB + SCC	3.01	3.29	12.31	6.77	7.44	21.46	15.81	22.42	27.16	11.89	20.31	21.23

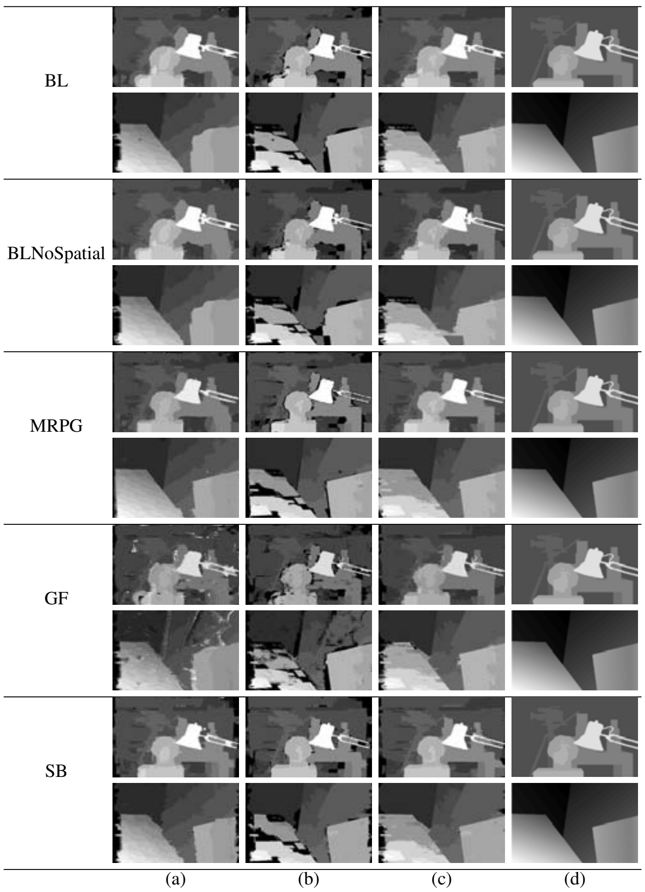


Fig. 5: Experimental results from proposed method. Raw disparity maps in (a), unknown disparities identified after Eq. 2 in (b), SCC final results in (c) and ground truth maps in (d).

V. Conclusion and Future Work

Support weighted windows have made important changes in the local stereo vision approach. Disparity maps can be yielded with high accuracy besides of maintaining edge preservation with a considerable quality.

In this work, we proposed a refinement disparity method based on a segmentation process and support weighted windows. The experimental results show that the proposed method is efficient and it can make some enhancements in disparity maps, as reducing the disparity error measure.

In the next phase of work, we want to compare this method with other post-processing techniques. In addition, we want to prepare new tests with the version 3 of the Middlebury Stereo Vision Evaluation.

Acknowledgment

The authors would like to thank IF Goiano (Instituto Federal Goiano) and FAPEG (Fundação de Apoio à Pesquisa de Goiás), process number 01/2018, for the financial support.

References

[1] 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.

[2] R. A. Hamzah and H. Ibrahim, “Literature survey on stereo vision disparity map algorithms,” *Journal of Sensors*, 2016.

[3] K.-J. Yoon and I.-S. Kweon, “Locally adaptive support-weight approach for visual correspondence search,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 2, June 2005, pp. 924–931 vol. 2.

[4] G. T. Laureano and M. S. V. de Paiva, “Disparities maps generation employing multi-resolution analysis and perceptual grouping,” in *2008 First Workshops on Image Processing Theory, Tools and Applications*, Nov 2008, pp. 1–6.

[5] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 504–511, Feb 2013.

[6] M. Gerrits and P. Bekaert, “Local stereo matching with segmentation-based outlier rejection,” in *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*. IEEE, 2006, pp. 66–66.

[7] A. Hosni, M. Bleyer, and M. Gelautz, “Secrets of adaptive support weight techniques for local stereo matching,” *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 620–632, 2013.

[8] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002.

[9] N. Ma, Y. Men, C. Men, and X. Li, “Accurate dense stereo matching based on image segmentation using an adaptive multi-cost approach,” *Symmetry*, vol. 8, no. 12, p. 159, 2016.

[10] F. Tombari, S. Mattoccia, and L. Di Stefano, “Segmentation-based adaptive support for accurate stereo correspondence,” *Advances in Image and Video Technology*, pp. 427–438, 2007.

[11] S. Mattoccia, “A locally global approach to stereo correspondence,” in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1763–1770.

[12] D. Chen, M. Ardabilian, and L. Chen, “A fast trilateral filter-based adaptive support weight method for stereo matching,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 730–743, May 2015.

[13] D. Scharstein and R. Szeliski, “Middlebury stereo visions.” [Online]. Available: http://vision.middlebury.edu/stereo/

[14] G. F. et al, “Integral images for block matching (demo).” [Online]. Available: http://demo.ipol.im/demo/57/