Stereo matching enhancement by statistical analysis and weighted functions

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*] and Cristhiane Gonçalves[7*]
[*]Federal University of Goiás, *Pixelab Laboratory*. Goiânia - GO, Brazil.
{fabrizzio[2], gustavo[3], ronaldocosta[6]}@inf.ufg.br, rafaeltp3[4]@gmail.com, cristhiane.goncalves[7]@ufg.br
[§]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 disparity refinement method to be applied in stereo matching algorithms. It consists of a segmentation process, statistical analysis of grouping areas and a support weighted function to find unknown disparities. We investigate the behavior of this method by comparing it with other post-processing techniques, as the left to right consistency-check. By comparing some of the most common refinement techniques, the experimental results show that or method achieved the lowest erros in non-weighted functions. Furthermore, through a qualitative evaluation, it is possible to note that our method reaches significant results, close to the ground truth maps.**

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

I. Introduction

A disparity map is a key component of a stereo vision system. It stores the differences between similar points, or pixels, from two or more images. Thus, in each point of a disparity map there is a value that represents the displacement of a point in a considered scene. In this field, the binocular approach is a common procedure. The idea is to simulate the human vision by using only two images. In most cases, these images are acquired through a stereo camera.

To perform a stereo algorithm, both images are labeled. One of them is the *reference image* and the other is the *target image*. The disparity map is built according to the reference image and the target image is used to find matching points among them. Thereby, it is possible to calculate the disparities.

In addition, the dissimilarities between points are evaluated by using cost functions. These functions hold a mathematical model that defines a way to compare points. For instance, the Sum of Absolute Differences (SAD) receives two values (or points) and performs the difference between them.

Most stereo vision methods follow the taxonomy proposed by [1]. It consists of four steps which defines a pipeline. The first step was defined as *matching cost*. The second is known as *cost aggregation*. The third is the *disparity selection* and the last is the *disparity refinement* step.

In this work, we propose a method to be applied in the fourth step from that pipeline. We explain the proposed approach and we prepare test cases to evaluate the methodology. Besides that, we compare it with other refinement techniques as left to right consistency-check. Experiments show that this technique is comparable with other post-processing strategies, in most cases with better results.

The outline of the paper is as follows. In Section II, we first discuss related work. In Section III, we present our algorithm for adjusting a disparity map. Finally, we present experimental results and conclusions in Section IV and V, respectively.

II. Related Work

Left to right consistency-check is a frequent refinement process that is applied in a raw disparity map. It consists of cross checking two or more disparity maps. For instance, to apply this methodology in a binocular approach, two disparity maps must be yielded, so each one of the image pair is used as a reference image, one at a time.

This consistency-check method deals with areas that are occluded. Points that are not visible may be detected and labeled as unknown disparities. After that, another technique can be used to fill in these occluded pixels. This method can improve a raw map by observing neighboring points. Therefore, if a disparity of a point is unknown, probably the neighboring points may tell what is.

Rhemann et al. [2] explains this method. When an occluded pixel is detected, it is assigned to the lowest disparity value of the spatially closest non-occluded pixels which lie on the same scanline (pixel row). However, they point out that this simple occlusion filling strategy can generate streak-like artifacts in the disparity map. Thus to remove them, while preserving the object boundaries, a weighted median filter can be applied to the filled pixels.

Hosni et al. [3] used this method with a simple modification. They observed that since occlusion occurs in the background of an image, the occluded pixel can be assigned to the minimum value from both disparity maps. According to them, this strategy also generates horizontal streaks in the disparity map and hence it demands post-processing on the filled in pixels.

This disparity map combination was also used by [4]. To improve the accuracy of their results, they calculated a depth image for both stereo images and combined them to eliminate some final artifacts. They assumed that artifact faults only occur in one of the two views, so they took the minimum of both disparity maps.

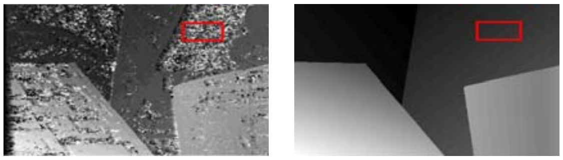
Apart from that method, other methodologies can be found. One of them was proposed by Mattocia el at. [5]. It is a powerful method that uses adaptive weights for classifying pixels based on geometric and photometric constraints. It takes a pair of images and a raw disparity map as input, then the plausibility of each point is evaluated by considering the relation among points in the same aggregating window, points between images and the original disparity.

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

In this brief section, some strategies of disparity enhancement were pointed out. It doesn't exclude other methodologies but it shows that the most discussed method in this literature is the left to right consistency-check. This is most likely due to its simplicity for implementing and ability to find occluded points with efficiency.

III. Proposed Approach

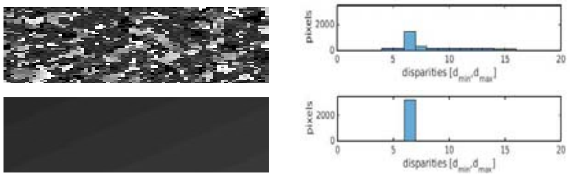
We start by analysing a raw disparity map. Fig. 1 shows a map that is very noisy in some parts of it. It was made by a simple cost aggregating (CA) methodology that can be called as *fixed window* (FW) method. It is the simplest CA strategy that uses an aggregating window and it is at the foundation 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.

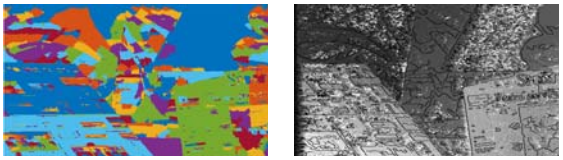


(a) (b)

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 [6] is widely employed. It was used to obtain great results in [4], [7] and [8]. 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 [9] 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 [5], [10], [11] among others and investigated in [3].

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

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

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

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

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

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

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

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

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 of 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.

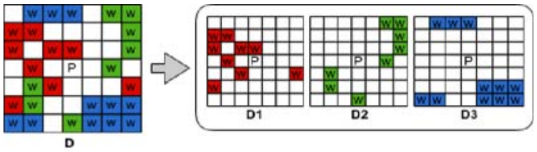


Fig. 4: Points in a disparity map D have their own weight w. Each segment is painted didactically (in red, green and blue). Weights in $D1$, $D2$ and $D3$ are summed separately. The best value is used to set the disparity in the principal point p.

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

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

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

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

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. In this work, we investigate the behavior of this method by comparing it with other post-processing techniques. Details are shown in the next section.

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 [12]. Each pair has its own ground truth that was used to evaluate the results.

We use the Root Mean Squared Error (RMSE) as a quantitative way to estimate the quality of the computed post-processing techniques. It computes the error between a computed disparity map $d_C(x, y)$ and its respective ground truth map $d_T(x, y)$, as given by

$$R = (\frac{1}{N} \sum_{(x,y)} |d_C(x, y) - d_T(x, y)|^2)^{\frac{1}{2}} \tag{10}$$

where N is the total number of pixels.

In the test case, 10 stereo vision methods were implemented. Thus, raw disparity maps from Fixed Window (FW), Shiftable Window (SW), Maximum Likelihood (MLMH), Large Occlusion (LC), Variable Window (VW), 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 test.

Furthermore, the methods discussed in Section II were also implemented. The left to right consistency-check with minimum horizontal neighbors, with median filter and with minimum disparities (CCMin, CCMedian and CCMinDisps, respectively), and the locally consistent method (LC). In the case of our method, it is referred to as *segment consistency-check* (SCC).

We used the SAD cost as a measure of stereo matching and a 39×39 support window for weighted functions. The color and spatial terms were set as $\gamma c = 23$ and $\gamma s = 14$ and these values were also set in SCC method. For the threshould in Eq. 2, it was set as $t = 1$. Other parameters were set as they were shown in their original reference.

For each combination (stereo method plus a post-processing technique) was calculated the RMSE. We used four image pairs (Tsukuba, Venus, Teddy and Cones), hence the average error between them was also calculated. Table I shows these results with the lowest errors in boldface.

We note that SCC method achieved important results, especially in non-weighted stereo methods. Thus, in the first five methods (from row 1 to 5) the error was decreased substantially in comparison to maps without the post-processing step (Raw column). However, the average errors were increased in the last five methods while CCMedian and CCMin obtained an improvement after their application in these weighted functions.

Fig. 5 shows some raw disparity maps from SW, LO, BL and MRPG methods. Besides, it shows their improvements made by CCMedian, LC and SCC methods. Through a qualitative evaluation, it is possible to note that SCC method reaches significant results, close to the ground truth maps.

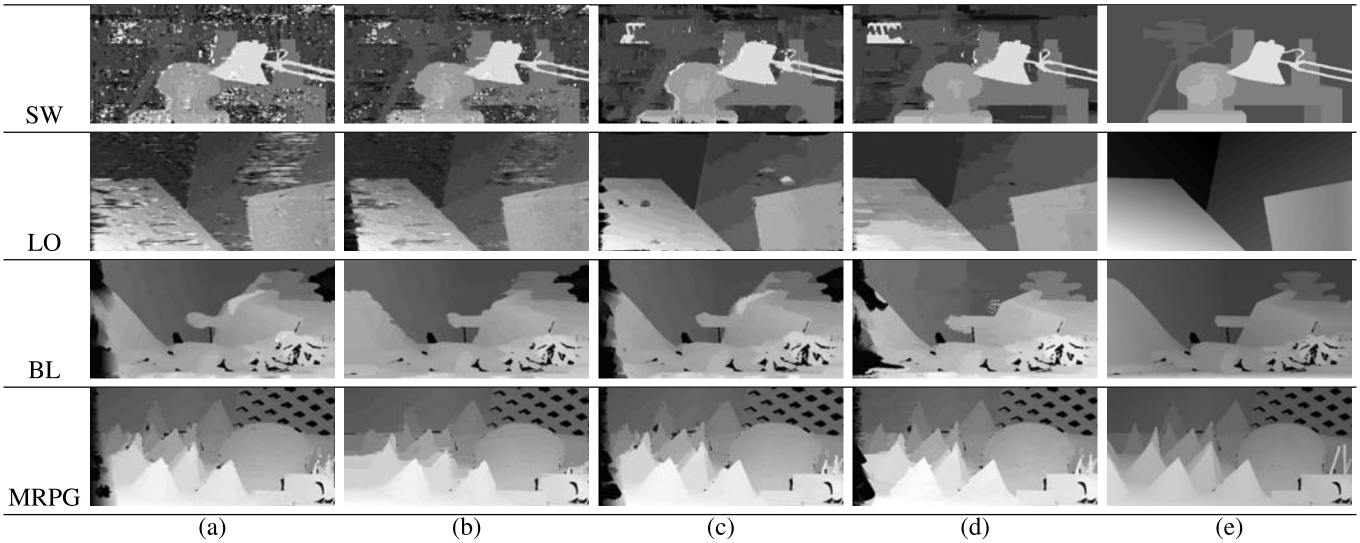


Fig. 5: Experimental results, (from top-down, Tsukuba with SW method, Venus with LO method, Teddy with BL method and Cones with MRPG method). Raw disparity maps in (a). CCMedian, LC and SCC outputs in (b), (c) and (d), respectively. Ground truth maps in (e).

TABLE I: Accuracy evaluation.

	Raw	CCMin	CCMedian	CCMinDisps	LC	SCC
FW [-]	8.15	6.51	**5.57**	8.15	7.44	6.79
SW [1]	8.37	8.86	8.53	8.65	7.81	**7.42**
MLMH [13]	14.74	10.20	10.07	15.29	12.77	**5.00**
LO [14]	6.94	9.26	9.21	8.56	7.01	**6.64**
VW [15]	8.33	7.13	**6.45**	8.46	7.69	7.27
BL [9]	7.20	4.36	**4.10**	7.17	7.39	7.39
BLNoSpatial [3]	7.21	4.38	**4.12**	7.21	7.46	7.36
MRPG [11]	6.39	4.19	**3.87**	6.51	6.78	6.47
GF [2]	7.62	4.66	**4.33**	7.55	7.56	7.61
SB [4]	7.30	4.15	**3.89**	7.24	7.47	7.38

V. Conclusion and Future Work

The proposed method is an effective post-processing technique. It improves disparity maps by using a segmentation process, statistical analysis of grouping areas and through a support weighted window to find unknown disparities.

By comparing the most common post-processing techniques, the experimental results showed that SSC method achieved some of the lowest errors in non-weighted functions. In the next study, we want to incorporate the left to right consistency-check in our method and to prepare a new 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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, "A maximum likelihood stereo algorithm," *Computer vision and image understanding*, vol. 63, no. 3, pp. 542–567, 1996.

[14] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 181–200, 1999.

[15] O. Veksler, "Fast variable window for stereo correspondence using integral images," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–I.