Cost Adaptive Window for Local Stereo Matching

J. Navarro and A. Buades
Dpt. Matemàtiques Informàtica, Universitat Illes Balears, Ctra Valldemossa km 7.5, Palma, Spain

Keywords: Stereovision, Depth Estimation, Block-matching, Adaptive Windows.

Abstract: We present a novel stereo block-matching algorithm which uses adaptive windows. The shape of the window is selected to minimize the matching cost. Such a window might be the less distorted by the disparity function and thus the optimal one for matching. Moreover, we introduce a coarse-to-fine strategy to limit the number of ambiguous matches and reduce the computational cost. The proposed approach performs as state of the art local matching methods.

1 INTRODUCTION

The goal of stereovision is to estimate the depth of the scene from at least two images taken from different viewpoints. The depth estimation is equivalent to computing the apparent motion of corresponding points in the two images. For an epipolar rectified image pair, all pixels have horizontal motion (which is called disparity) and the problem is then reduced into a 1D correspondence problem.

Over last years, several approaches have been proposed to solve the stereo matching problem. The strategies can be divided into local and global methods (Scharstein and Szeliski, 2002). Local methods compute the disparity d of a point (x, y) by means of a block-matching (also called area-based) approach in which a small window or patch around (x, y) in the left image is compared with windows in the same epipolar line in the right image. The comparison is done by assigning a matching cost c to each candidate window in the second image. The global methods overcome the main limitation of block based methods, that is, the non presence of enough distinctive information in the block. For global methods, the disparity estimation is formulated by means of an optimization problem where the solution is constrained to satisfy some smoothness assumption. These methods, being very similar to optical flow methods, mainly differ in the energy minimization method being used, belief propagation (Sun et al., 2003), graph cuts (Kolmogorov and Zabih, 2001), etc. Global methods depend on additional parameters difficult to fix in general, its value being different for each stereo pair.

In order to reliably match a block, the depth should vary as less as possible inside the block. Otherwise, the block might be distorted by the effect of the disparity function and its retrieval in the second image not be an easy task. Block matching methods tend to identify depth discontinuities with image color ones, and adapt the shape of the window to the color of the image (Patricio et al., 2004; Yoon and Kweon, 2006; Wang et al., 2006; Rhemann et al., 2011). This solution might be effective for scenes with uniform color objects being at different depth planes, however this is not the case for a general scenario with textured objects and slanted surfaces.

We propose to adapt the shape of the window to the unknown disparity function instead of the image color. Windows sharing the same depth for all pixels will be less distorted by the application of the disparity and therefore will be matched with minimal cost. Similarly, the authors in (Buades and Facciolo, 2015; Hirschmüller et al., 2002) proposed a similar approach but limiting the choice of the matching window to belong to a small pre existing set, containing mainly directional and corner windows. As we will see, the choice of a correct window is not straighforward and some regularity on the shape of the window might be demanded.

The proposed approach is able to deal with depth discontinuities and slanted surfaces. Compared to state-of-the-art, the proposed approach is able to precisely identify depth discontinuities avoiding the well known fattening problem (Blanchet et al., 2011). The algorithm is embedded into a coarse-to-fine strategy in which the disparity computed at a coarser scale is used to restrict the search disparity range at finer scales reducing both the match ambiguity and com-

Navarro J. and Buades A.
Cost Adaptive Window for Local Stereo Matching.
DOI: 10.5220/0006100503690376
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 369-376
ISBN: 978-989-758-227-1

putational time.

This paper is organized as follows. In Section 2 we present the state-of-the-art of local methods for disparity estimation. In Section 3 it is explained the new block-matching algorithm with adaptive windows. Finally, Section 4 shows the performance of the proposed method by means of a comparison with state-of-the-art approaches.

2 STATE OF THE ART

Local methods mainly differ in the choice of the matching cost and the size and shape of the matching window. The most common costs are the sum of squared differences, the normalized cross correlation (Hannah, 1974), the summed normalized cross correlation (Einecke and Eggert, 2010), the mutual information (Viola and Wells III, 1997) and the census transform (Zabih and Woodfill, 1994), see (Hirschmüller and Scharstein, 2009) for a review. While the choice of a different cost might be important in order to be robust to local differences in color, noise, shadows or transparencies, the shape and size of the window turn to be the most important in order to overcome the effect of disparity in the image.

Block-matching algorithms typically assume that the depth is the same for all pixels inside the matching window. This assumption may not hold for slanted surfaces or near depth discontinuities unless the shape of the window is locally adapted. Kanade et al. (Kanade and Okutomi, 1994) were the first to address this problem. The authors proposed the use of rectangular adaptive windows whose shape and size is selected locally at each pixel by looking the differences between gray level values. Approaches with fixed size and squared windows were presented in (Fusiello et al., 1997; Kang et al., 2001) where it is performed the correlation with different windows containing the reference pixel and then it is taken the disparity giving the smallest cost, arguing that a window yielding the smallest error is more likely to correspond to a constant depth region. There are also methods that base the shape and size of the correlation window on an image segmentation (Gerrits and Bekaert, 2006; Wang and Zheng, 2008). Other approaches, instead of adapting the window make use of varying weights for the pixels inside the fixed window (Yoon and Kweon, 2006; Wang et al., 2006; Rhemann et al., 2011). See (Hosni et al., 2013) for an extensive review. All these methods identify differences in pixels' gray level with differences in pixels' depth, which is not always the case.

All the methods mentioned so far are expected to cope with depth discontinuities but not with slanted surfaces. However, proposals have arisen with the aim of handling the presence of slanted surfaces in the scene. There are approaches trying to estimate depth by means of planes at each region (Lu et al., 2013; Bleyer et al., 2011). Other approaches cope with slanted surfaces by selecting the most appropriate window at each pixel depending on cost, using a pre existent set of windows. In (Hirschmüller et al., 2002) the shape of the window was adapted by dividing the correlation window into sub-windows and selecting the ones yielding the minimum matching cost. The recent method presented in (Buades and Facciolo, 2015) consists of a parameter-less approach in which multiple elongated windows with different orientations are tested in the matching process. At each point, the one providing minimum matching cost is selected. The authors also introduce a set of validation criteria in order to provide a mask of invalid or ambiguous matches. Furthermore, they use a coarse-to-fine strategy in which disparity computed at a coarser scale is used to restrict the search disparity range at the current scale.

One of the main problems of local methods is the errors that can be produced in non-textured regions. The lack of information in this patches lead to ambiguities in the matching process. The authors in (Manduchi and Tomasi, 1999) target this problem by computing the distinctiveness of a pixel as the dissimilarity in color between the pixel and the most similar other point in the search window. The recent approach proposed in (Sabater et al., 2012) measures reliability from the number of false alarms to each match and matches whose patches are slightly different are discarded. Also, they use the distinctiveness measure presented in (Manduchi and Tomasi, 1999).

Finally, the sampling of the disparity space is also an important question since it can produce incorrect matches. In (Birchfield and Tomasi, 1998) the authors proposed a matching cost based on the SSD that is insensitive to the sampling of the correlation window. However, in (Buades and Facciolo, 2015) the authors show that similar results can be obtained with a subpixel SSD cost.

3 BLOCK-MATCHING WITH ADAPTIVE WINDOWS

Block-matching methods assume that disparity does not change drastically inside the window. A proper choice of the window shape is important in order to satisfy such a condition. This optimum shape has to be adapted locally for each pixel.

The optimal window cannot be known a priori, however the analytical study of correlation in (Blanchet et al., 2011) permits to characterize the pixel minimizing such a cost for a fixed window. Following (Blanchet et al., 2011), the optimum match writes as a weighted average of disparities of pixels in the window, being this average weighted by the gradient of the image. Such an average would coincide with the true disparity when the disparity is constant and the cost be minimal.

We propose to select at each pixel a matching window sharing the same depth, indirectly by choosing the one matching with minimal cost. This window has to be adapted independently for each pixel and for each possible disparity, then the disparity and window with minimal cost have to be chosen. In order to balance the shape of the window around the reference pixel we penalize the distance of the barycenter of the window to the reference pixel. We do not penalize the distance of the chosen pixels to the reference pixel since we allow for elongated windows. Although marginally, in some cases it may be necessary to take into account the intensity or color of the pixels belonging to the selected window. Thereby, we select only pixels having a similar color to the reference pixel.

The selection of the adapted window $W_{\mathbf{p}}$ for a certain pixel $\mathbf{p}$ and disparity d writes as a minimization

$$\mathcal{C}(\mathbf{p},d) = \min_{W_{\mathbf{p}}} \sum_{\mathbf{p}_i \in W_{\mathbf{p}}} c(\mathbf{p}_i,d) + \beta \|\mathbf{p} - \mathbf{b}_{\mathbf{W_p}}\|^2 + \lambda \sum_{\mathbf{p}_i \in W_{\mathbf{p}}} (I_1(\mathbf{p}) - I_1(\mathbf{p}_i))^2, \quad (1)$$

being $c(\mathbf{p_i},d)$ the cost of assigning disparity d to the pixel $\mathbf{p_i}$, $\mathbf{b}_{\mathbf{W_p}}$ the barycenter of the chosen window and $\lambda,\beta > 0$.

Then, for each pixel $\mathbf{p}$ we compute the disparity $D(\mathbf{p})$ as

$$D(\mathbf{p}) = \arg\min_{d} \mathcal{C}(\mathbf{p},d). \quad (2)$$

3.1 Adaptive Window Selection

Since not all possible window configurations can be tested, we perform a greedy algorithm making use of a cost volume structure, similarly to (Rhemann et al., 2011). First, in order to be robust we use 3×3 squared windows for building the cost volume $c(\mathbf{p},d)$ as the three dimensional array which stores the zero-mean sum of squared differences (ZSSD) cost (Hirschmüller and Scharstein, 2009) for choosing disparity d at pixel $\mathbf{p} = (x,y)$. The ZSSD cost removes the average intensity of the window rendering the comparison independent of the mean intensity:

$$ZSSD(\mathbf{p},\mathbf{q}) := \frac{1}{|B_r|}\sum_{\mathbf{t}\in B_r}\left| I_1(\mathbf{p}+\mathbf{t}) - I_2(\mathbf{q}+\mathbf{t}) - \overline{I_1|_{\mathbf{p}+B_r}} + \overline{I_2|_{\mathbf{q}+B_r}}\right|^2, \quad (3)$$

where $\mathbf{q} = \mathbf{p} + (d,0)^\top$, B_r denotes the matching window and $\overline{I_1|_{\mathbf{p}+B_r}}$ is the mean of pixel intensities in image I_1 inside the window with reference pixel $\mathbf{p}$.

Then, we select at each pixel $\mathbf{p}$ the window $W_{\mathbf{p}}$ that minimizes (1). This window is composed by neighboring 3×3 patches of $\mathbf{p}$. For a given n, we select the $n-1$ neighboring patches that minimize such an energy. These patches are selected with an iterative process, selecting at each step the patch yielding mininum cost and being connected to the already selected patches. That is, by construction the window is connected. Moreover, in order to control the elongation of the optimal window, we introduce a parameter M to limit the distance between the reference pixel $\mathbf{p}$ and the center pixel of the joined patch.

Let $\mathbf{b}_{\mathbf{1},\ldots,\mathbf{k}}$ be the barycenter of points $\mathbf{p},\mathbf{p_1},\ldots,\mathbf{p_k}$. At iteration k we add the neighboring patch centered at $\mathbf{p_k}$ such that

$$\mathbf{p_k} = \arg\min_{\substack{\mathbf{p_{k'}} \\ \mathbf{p_{k'}}\in N_{1,\ldots,k-1} \\ \|\mathbf{p}-\mathbf{p_{k'}}\|\le M}} c(\mathbf{p_{k'}},d) + \beta\|\mathbf{p} - \mathbf{b}_{\mathbf{1},\ldots,\mathbf{k'}}\|^2 + \lambda\left(\overline{I_1|_{\mathbf{p}+B_3}} - \overline{I_1|_{\mathbf{p_{k'}}+B_3}}\right)^2 \quad (4)$$

where $\mathbf{b}_{\mathbf{1},\ldots,\mathbf{k'}}$ is the updated barycenter taking into account $\mathbf{p_{k'}}$ and $N_{1,\ldots,k-1}$ the set of connected neighboring patches to the already selected ones. We repeat this process until we add $n-1$ patches centered at $\mathbf{p_1},\ldots,\mathbf{p_{n-1}}$.

The difference between the average color of the 3×3 candidate patches and the reference one is used to penalize windows across image edges. The use of the average value is more robust than comparing only the value of pixel p.

Now, the cost of assigning disparity d to the pixel $\mathbf{p}$ with the selected window is given by

$$A(\{\mathbf{p},\mathbf{p_1},\ldots,\mathbf{p_{n-1}}\},d) = \sum_{i=0}^{n-1} c(\mathbf{p}_i,d) + \beta\|\mathbf{p}-\mathbf{b}_{\mathbf{1},\ldots,\mathbf{n-1}}\|^2 + \lambda\sum_{i=1}^{n-1}\left(\overline{I_1|_{\mathbf{p}+B_3}} - \overline{I_1|_{\mathbf{p}_i+B_3}}\right)^2, \quad (5)$$

being $\mathbf{b}_{\mathbf{1},\ldots,\mathbf{n-1}}$ the barycenter of the chosen window. For notation purposes, notice that we have denoted $\mathbf{p}$ as $\mathbf{p_0}$ to introduce it into the sum. The shape resulting

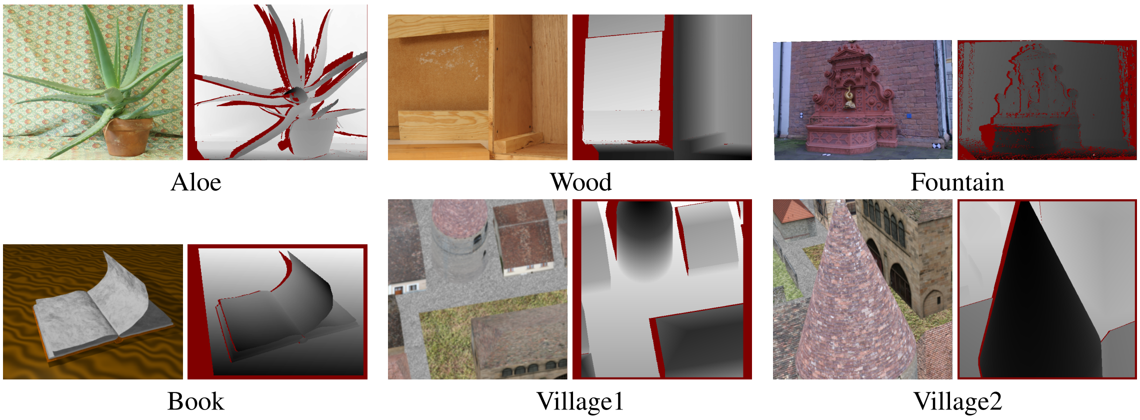

Figure 1: Images used in the experiments section to evaluate the performance of our method. We show the left image of the stereo pair and the ground truth disparity with the true occlusion mask overimposed.

from accumulating the smallest matching costs coincides with the one having the disparity function most uniform.

The volume of costs $A(\{\mathbf{p}, \mathbf{p_1}, \ldots, \mathbf{p_{n-1}}\}, d)$ turns to be an approximation of the cost $\mathcal{C}(\mathbf{p}, d)$. Then, for each pixel $\mathbf{p}$ we finally compute the disparity $D(\mathbf{p})$ as

$$D(\mathbf{p}) = \arg\min_{d} A(\{\mathbf{p}, \mathbf{p_1}, \ldots, \mathbf{p_{n-1}}\}, d). \qquad (6)$$

3.2 Multi-scale Approach

As in (Buades and Facciolo, 2015), our approach is embedded into a coarse-to-fine strategy. This permits to reduce computational cost and match ambiguity. A large search range may lead to mismatches as increases the possibility of matching with a repetitive pattern. For this reason, at each scale we adapt the search range locally at each pixel by looking at the minimum and maximum disparity values of a neighborhood obtained at the previous coarser scale. With this step we compute the images *Dmin* and *Dmax* which determine respectively the minimum and maximum values of the disparity range at each pixel. Each level of the pyramid is obtained by a convolution with a Gaussian kernel with standard deviation $\sigma = 1.2$ and subsampling by a factor of two from the initial stereo pair. The disparity computed at coarser scales is upsampled by bicubic spline interpolation.

3.3 Detection of Invalid Matches

In order to reject possible incorrect matches we use the common left-right consistency check. Let D_L and D_R be the left-based and right-based disparities. This is,

$$I_1(\mathbf{p}) = I_2(\mathbf{p} + D_L(\mathbf{p})) \quad \text{and} \quad I_1(\mathbf{p} + D_R(\mathbf{p})) = I_2(\mathbf{p}). \qquad (7)$$

Then, a match is rejected if D_R does not coincide with the inverse mapping of D_L, i. e. when

$$|D_R(\mathbf{p} + D_L(\mathbf{p})) + D_L(\mathbf{p})| > \varepsilon \qquad (8)$$

with $\varepsilon > 0$. In practice it is set to $\varepsilon = 1$.

4 EXPERIMENTS

For evaluating the proposed multi-scale adaptive window (MSAW) approach we use two databases: the set of images presented in Figure 1 and the new Middlebury stereo benchmark version 3 (Scharstein and Hirschmüller, 2014).

In this section we compare the proposed algorithm with several state-of-the-art methods. The Adaptive Support Weights (ASW) approach (Yoon and Kweon, 2006) uses bilateral weights with the aim of assigning large weights to close pixels having a similar color to the reference one, expecting to use pixels only from the same physical object. A window size of 35×35 is used. This method provides only pixel precision.

The Cost-Volume Filtering (CVF) method (Rhemann et al., 2011) builds a volume $c'(\mathbf{p}, d)$ measuring the cost for selecting the disparity d at pixel $\mathbf{p}$. Then, the guided filter (He et al., 2010) is used to filter the volume using the left image as a guidance. The patch size used in this approach is 19×19 and the method provides pixel precision.

Finally, the Multi-Scale and Multi-Window (MSMW) method (Buades and Facciolo, 2015) consists of a multi-scale approach which uses multiple elongated windows with different orientations in order to cope with slanted surfaces. We run the algorithm with three scales, a window size of 5×5 and $1/4$ of disparity precision. For this algorithm we only apply the matching strategy and not the posterior validation criteria defined in the method. We do

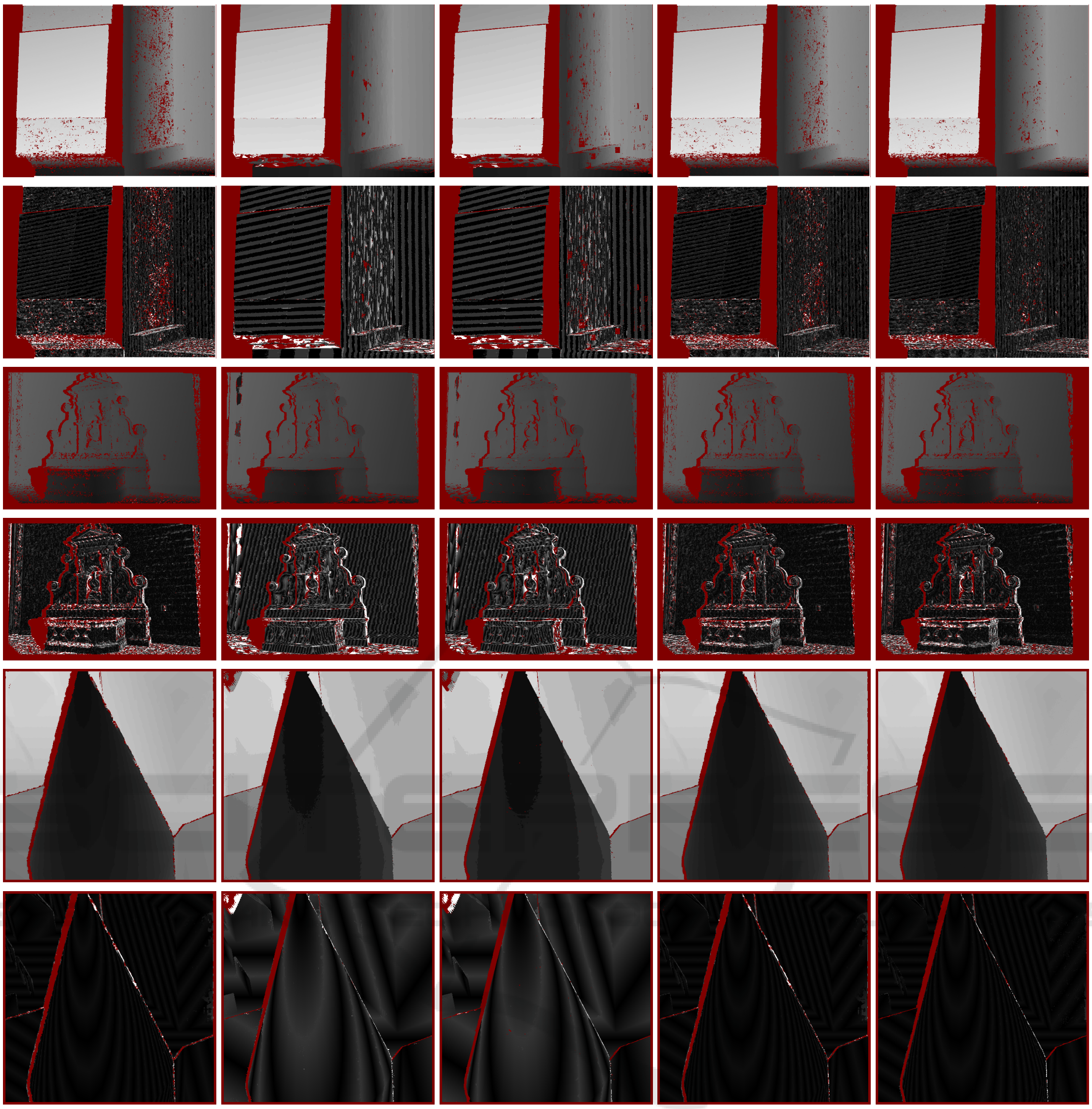

Figure 2: Disparity estimation and error, both with ground truth mask and left-right mask overimposed, for Aloe, Wood, Fountain and Village2 images of the proposed method (MSAW) compared with a classic block matching with 5×5 squared windows (Simple), Adaptive Support Weights (ASW) (Yoon and Kweon, 2006), Cost-Volume Filtering (CVF) (Rhemann et al., 2011) and Multi-Scale and Multi-Window algorithm (MSMW) (Buades and Facciolo, 2015). Image error between computed disparities and ground truth is displayed in range [0, 2.5].

this in order to fairly compare all algorithms, since all block-matching algorithms could benefit of such criteria. For completeness in the comparison, we include a matching algorithm with a fixed squared 5×5 window with $1/4$ of precision. We will refer to this algorithm as Simple.

Regarding our method, we use a fixed set of parameters in all the experiments. We take $n = 25$ as the number of patches to be joined. The maximum allowed distance when joining a new pixel is $M = 4$, the penalization parameters are $\lambda = 0.05$ and $\beta = 0.5$, we use $1/4$ of precision in the disparity map and we perform a total of three scales in the pyramidal scheme. For all methods we use only the left-right consistency check as validation. Figure 1 displays the image pairs used for testing in this section. For each pair we dispose of a different initial disparity range: $[-150.50, -21.50]$ for the Aloe pair, $[-108, -28]$ for Wood, $[-84, 200]$ for Fountain, $[-60, -21]$ for Book, $[-22, 23]$ for Village1 and the range $[-9, 10]$ for the Village2 stereo pair.

Figure 2 visually compares the described methods

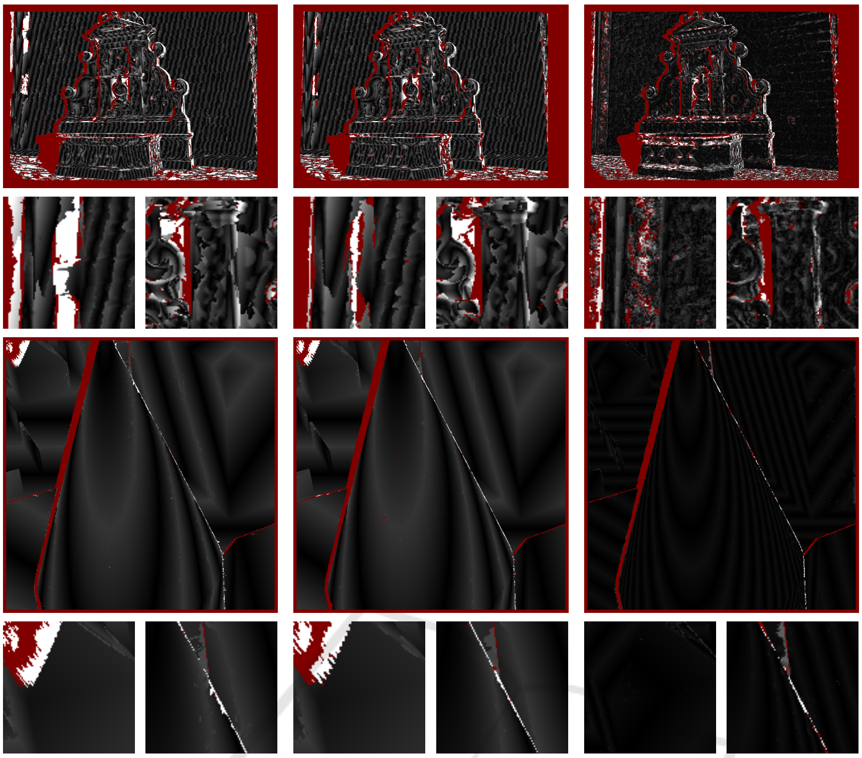

Figure 3: Comparison of the proposed method (MSAW) with Adaptive Support Weights (ASW) (Yoon and Kweon, 2006) and Cost-Volume Filtering (CVF) (Rhemann et al., 2011). It is shown the error with the ground truth mask and left-right mask overimposed for Fountain and Village2 pairs, displayed in range [0, 2.5].

Table 1: Comparison of our method (MSAW) with a classic block matching with 5×5 squared windows (Simple), Adaptive Support Weights (ASW) (Yoon and Kweon, 2006), Cost-Volume Filtering (CVF) (Rhemann et al., 2011) and the Multi-Scale and Multi-Window algorithm (MSMW) (Buades and Facciolo, 2015). We show the density (D), the percentage of pixels with an error above one pixel (E1) and above three pixels (E3), being all values computed for all the pixels.

	Simple			ASW			CVF			MSMW		
image	D	E1	E3	D	E1	E3	D	E1	E3	D	E1	E3
Aloe	82.16	4.08	2.26	86.33	3.59	1.85	84.34	3.08	1.61	83.33	2.98	1.70
Wood	79.41	4.34	1.98	84.74	3.63	1.10	82.91	2.71	0.87	83.15	2.63	0.90
Fountain	81.99	12.26	4.66	87.14	13.03	4.9	85.19	10.49	3.21	84.98	10.26	3.24
Book	87.03	6.02	1.79	89.38	3.59	0.15	87.57	**3.41**	**0.14**	87.60	5.13	1.25
Village1	92.71	2.62	1.46	94.36	2.92	0.72	93.81	2.22	0.56	93.01	2.07	1.10
Village2	97.16	0.61	0.42	97.57	0.65	0.61	97.48	0.61	0.61	97.31	0.47	0.35
average	86.74	4.99	2.10	89.92	4.57	1.56	88.55	3.75	1.17	88.23	3.92	1.42

	MSAW		
image	D	E1	E3
Aloe	84.49	**2.58**	**1.29**
Wood	84.29	**2.17**	**0.44**
Fountain	85.25	**9.73**	**2.52**
Book	84.79	4.85	0.88
Village1	93.65	**1.24**	**0.53**
Village2	97.82	**0.26**	**0.24**
average	88.38	**3.47**	**0.98**

for images Aloe, Wood, Fountain and Village2. In the figure we show the estimated disparity map and its error with the ground truth displayed in the range $[0, 2.5]$. We over-impose the true occlusion mask and the one resulting from the left-right consistency check. Near the contour of the leaves of the Aloe and in the contours of the buildings in Village2 we can appreciate how our method is the one that yields low-

Table 2: Evaluation on Middlebury online benchmark version 3. Comparison between our method and SGM (Hirschmüller, 2008), SNCC (Einecke and Eggert, 2010), IDR (Kowalczuk et al., 2013), Cens5 (Hirschmüller et al., 2002), TMAP (Psota et al., 2015), CVF (Rhemann et al., 2011), ASW (Yoon and Kweon, 2006) and MSMW (Buades and Facciolo, 2015). The numbers represent the proposed weighted average for the training dataset over non occluded pixels.

method	resolution	density (%)	bad 0.5 (%)	bad 1.0 (%)	bad 2.0 (%)
SGM	H	84.04	38.5	17.7	7.66
SNCC	H	71.29	29.8	13.4	6.14
IDR	H	76.10	31.9	12.8	**4.67**
Cens5	H	76.38	35.3	17.5	8.37
TMAP	H	69.80	37.3	15.4	5.97
CVF	H	62.62	34.13	17.13	12.03
ASW	H	66.05	42.19	26.78	21.73
MSMW	H	66.90	28.37	14.52	9.08
MSAW	Q	80.50	**19.48**	**11.10**	7.12

est errors near depth discontinuities. Larger errors of AWS and CVF are due to the different sampling of the disparity space. In general AWS and CVF behave similar, in Figure 3 we can see the comparison of our method and these two methods with cropped regions of Fountain and Village2 images. The regions of large error that we see in both images for AWS and CVF disappear in our result. At the top-left part of the Village2 image there is a slanted roof in which disparity is well estimated by our method while the two others fail.

In Table 1 we show a quantitative comparison with the mentioned methods. In both tables we show the density (D) of each method obtained from the left-right consistency check, the percentage of pixels with an error above one pixel (E1) and with an error above three pixels (E3). We use all the pixels (occluded and non occluded) in the evaluation. The proposed algorithm is the one achieving in general lower errors, obtaining the lowest average in both cases, all and non occluded pixels.

Finally, we apply our method MSAW to the new Middlebudy benchmark version 3 (http://vision.middlebury.edu/stereo/eval3/). In Table 2 we show the results provided by other published approaches that are in the highest positions of the ranking (September 2016): SGM (Hirschmüller, 2008), SNCC (Einecke and Eggert, 2010), IDR (Kowalczuk et al., 2013), Cens5 (Hirschmüller et al., 2002) and TMAP (Psota et al., 2015). Additionally, we also include the results given by the methods used in the previous comparison: CVF (Rhemann et al., 2011), AWS (Yoon and Kweon, 2006) and MSMW (Buades and Facciolo, 2015). Our method gives a considerable number of estimated pixels while yielding low errors.

5 CONCLUSIONS

We proposed a new local matching algorithm to compute the disparity from a stereo rectified image pair. Our method adapts for each pixel the shape of the window in order to select the one with minimal matching cost. This window being the less distorted by the disparity coincides with the one for which the depth varies the least and then the optimal one for local matching.

The proposed algorithm makes use of a cost volume for selecting the minimal cost window. Additional criteria is used to balance the shape of the window and avoid windows containing more than one physical object. Moreover, we use a pyramidal strategy in which the search disparity range is restricted at each scale reducing the match ambiguity and the computational time.

The experiments show how the proposed method outperforms state-of-the-art local matching methods being able to deal with depth discontinuities and slanted surfaces.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support by Ministerio de Economia y Competitividad under grant TIN2014-53772-R and CNES research and technology project DAJ/AR/IB/16-10117037.

REFERENCES

Birchfield, S. and Tomasi, C. (1998). A pixel dissimilarity measure that is insensitive to image sampling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(4):401–406.

Blanchet, G., Buades, A., Coll, B., Morel, J.-M., and Rougé, B. (2011). Fattening free block matching. *Journal of mathematical imaging and vision*, 41(1-2):109–121.

Bleyer, M., Rhemann, C., and Rother, C. (2011). Patchmatch stereo-stereo matching with slanted support windows. In *BMVC*, volume 11, pages 1–11.

Buades, A. and Facciolo, G. (2015). Reliable multiscale and multiwindow stereo matching. *SIAM Journal on Imaging Sciences*, 8(2):888–915.

Einecke, N. and Eggert, J. (2010). A two-stage correlation method for stereoscopic depth estimation. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 227–234. IEEE.

Fusiello, A., Roberto, V., and Trucco, E. (1997). Efficient stereo with multiple windowing. In *cvpr*, page 858. IEEE.

Gerrits, M. and Bekaert, P. (2006). Local stereo matching with segmentation-based outlier rejection. In *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*, pages 66–66. IEEE.

Hannah, M. J. (1974). Computer matching of areas in stereo images. Technical report, DTIC Document.

He, K., Sun, J., and Tang, X. (2010). Guided image filtering. In *Computer Vision–ECCV 2010*, pages 1–14. Springer.

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341.

Hirschmüller, H., Innocent, P. R., and Garibaldi, J. (2002). Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1-3):229–246.

Hirschmüller, H. and Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(9):1582–1599.

Hosni, A., Bleyer, M., and Gelautz, M. (2013). Secrets of adaptive support weight techniques for local stereo matching. *Computer Vision and Image Understanding*, 117(6):620–632.

Kanade, T. and Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(9):920–932.

Kang, S. B., Szeliski, R., and Chai, J. (2001). Handling occlusions in dense multi-view stereo. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–103. IEEE.

Kolmogorov, V. and Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 508–515. IEEE.

Kowalczuk, J., Psota, E. T., and Perez, L. C. (2013). Real-time stereo matching on cuda using an iterative refinement method for adaptive support-weight correspondences. *IEEE transactions on circuits and systems for video technology*, 23(1):94–104.

Lu, J., Yang, H., Min, D., and Do, M. (2013). Patch match filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1854–1861.

Manduchi, R. and Tomasi, C. (1999). Distinctiveness maps for image matching. In *iciap*, page 26. IEEE.

Patricio, M. P., Cabestaing, F., Colot, O., and Bonnet, P. (2004). A similarity-based adaptive neighborhood method for correlation-based stereo matching. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 2, pages 1341–1344. IEEE.

Psota, E. T., Kowalczuk, J., Mittek, M., and Perez, L. C. (2015). Map disparity estimation using hidden markov trees. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2219–2227.

Rhemann, C., Hosni, A., Bleyer, M., Rother, C., and Gelautz, M. (2011). Fast cost-volume filtering for visual correspondence and beyond. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3017–3024. IEEE.

Sabater, N., Almansa, A., and Morel, J.-M. (2012). Meaningful matches in stereovision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(5):930–942.

Scharstein, D. and Hirschmüller, H. (2014). Middlebury stereo evaluation version 3. *http://vision.middlebury.edu/stereo/eval3/*.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42.

Sun, J., Zheng, N.-N., and Shum, H.-Y. (2003). Stereo matching using belief propagation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(7):787–800.

Viola, P. and Wells III, W. M. (1997). Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154.

Wang, L., Liao, M., Gong, M., Yang, R., and Nister, D. (2006). High-quality real-time stereo using adaptive cost aggregation and dynamic programming. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 798–805. IEEE.

Wang, Z.-F. and Zheng, Z.-G. (2008). A region based stereo matching algorithm using cooperative optimization. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Yoon, K.-J. and Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):650–656.

Zabih, R. and Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In *Computer VisionECCV'94*, pages 151–158. Springer.