A SIMILARITY-BASED ADAPTIVE NEIGHBORHOOD METHOD FOR CORRELATION-BASED STEREO MATCHING

Madaín Pérez Patricio, François Cabestaing, Olivier Colot, Pierre Bonnet

LAGIS, Laboratoire d'Automatique, Génie Informatique & Signal, UMR CNRS 8146
USTL, Cité Scientifique, 59655 Villeneuve d'Ascq CEDEX, France.
mp@i3d.univ-lille1.fr, fc@ieee.org, {olivier.colot;pierre.bonnet}@univ-lille1.fr

ABSTRACT

In this paper, we present a new method for dense stereo matching. In area-based methods, the similarity between one pixel of the left image and one pixel of the right image is measured using a correlation index computed on neighborhoods of these pixels. In our method, the neighbor pixels not similar to the center one are excluded when computing the correlation index, which corresponds to adjusting the equivalent size and shape of the correlation neighborhood. Our algorithm yields a precise estimation of the disparity in non-textured areas, while avoiding undesired smoothing at discontinuities . This method is suitable for real-time implantation using specialized hardware. We demonstrate and discuss performances using synthetic stereo pairs.

Keywords: adaptive neighborhood; dense stereovision; video-rate processing.

1. INTRODUCTION

Stereo algorithms allow the depth information to be determined in a scene by combining two images of this scene taken at the same time from slightly different viewpoints. The correspondence problem consists in defining conjugate pairs of pixels, one from each image, that correspond to the same point of the scene. Then, a standard triangulation technique yields the depth of this point. A sparse correspondence method yields the depth information for some specific points of the scene while a dense solution yields the depth of all the visible points of the scene. For applications like mobile robotics, a dense depth map and a video rate processing of the images are often required.

We assume that both images have been rectified, i.e. that the epipolar lines are the raster lines. In area-based methods, a neighborhood of a reference pixel in one of the images is compared to a similar neighborhood of every pixel in the same raster line of the other image. Usually, neighborhoods are rectangular windows centered on the pixels. First, a similarity index is determined for each candidate pixel, based on the contents of the two neighborhoods. Then, the maximum (or minimum) value of this similarity index defines the most relevant candidate in the second image and the shift is retained as the disparity value. Most area-based approaches use a correlation index as a similarity measure for determining the best match.

Selecting the size of the correlation window is a difficult task. With a large window, used to reveal maxima of correlation in non-textured areas, edges are blurred and small details or small objects are removed. On the other hand, with a small window, the correlation index is a measure very sensitive to noise. Several methods have been proposed to improve the matching efficiency at depth discontinuities. Kanade and Okutomi have proposed an adaptive neighborhood method [1], in which they iteratively change the neighborhood size and shape according to the local variation of the intensity and current depth estimates. However, the algorithm is computationally expensive [2].

To simplify the adaptive methods, efficient multiple windows methods have been proposed. Roberto et al. have described a symmetric, multi-window (SMW) algorithm [3]. They compute SSD indices on nine windows, and retain the one yielding the lowest value of the SSD. Hirschmüller uses a central window surrounded by several support windows [4]. The correlation indices of the best support windows, i.e. the lowest values, are added to the index computed on the central window. To make the value of the similarity index less sensitive to outliers, Zabih and Woodfill have used non-parametric local transforms of the neighborhood data [5]. Zhang and Kambhamettu have proposed matching with a segmentation-based cooperation technique [6].

In this paper, we describe an algorithm in which the size and shape of the correlation neighborhood are adjusted according to its content. Each pixel of a fixed size rectangular window is included or not in the adaptive neighborhood according to its similarity with the center pixel. The neighborhood is not the rectangular window like in many other methods, but only a subset of it. The paper is organized as

The authors express their gratitude to COSNET-MEXICO for the COSNET scholarship 2001196P, and acknowledge the financial support of the French Nord-Pas de Calais Regional Council.

follows : section 2 presents our algorithm; experimental results with two synthetic stereo pair are reported in section 3 and section 4 concludes this paper.

2. SIMILARITY-BASED ADAPTIVE NEIGHBORHOOD METHOD

Thereafter, $P_l(x, y)$ (resp. $P_r(x, y)$) denotes the pixel with coordinates (x, y) in the left (resp. right) image, and $I_l(x, y)$ (resp. $I_r(x, y)$) its grey level. $d(x, y)$ (resp. $d_r(x, y)$) is the disparity for the pixel of the left (resp. right) image with coordinates (x, y). $w \times w$ is the size of the square correlation window.

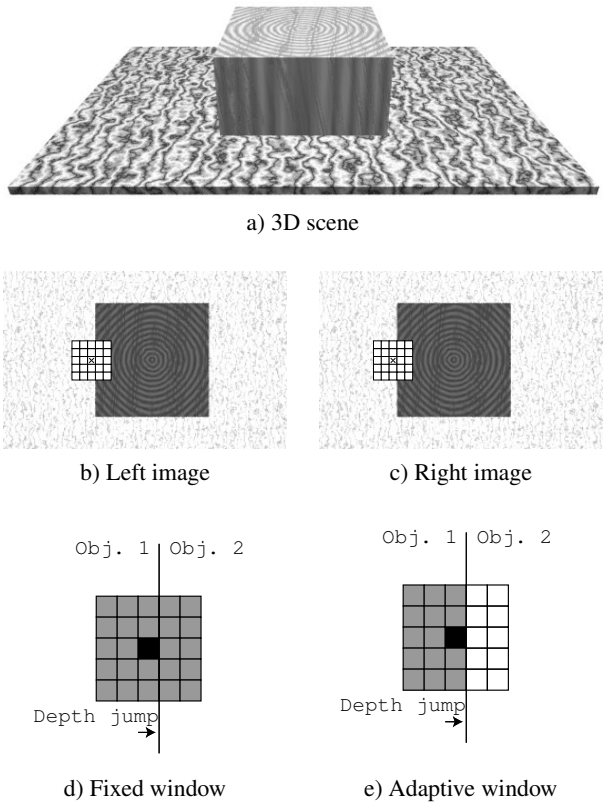
If the correlation window covers a region of the image in which the depth of scene points varies, then standard area-based matching techniques fail. For example, consider the case of the scene shown in figure 1(a), where two textured objects with different depths are seen by the cameras. The corresponding left and right images are shown in figures 1(b) and 1(c). The pixels within the small overlapped windows are the projections of points of the two objects. When the correlation index is computed using all the pixels of this window (cf. figure 1(d)), the averaging effect yields errors on the estimated disparity. On the other hand, the adaptive window shown in figure 1(e) includes only the leftmost pixels, that are the projections of points of the same object. With this window, disparity estimation is more precise.

In our Similarity Based Adaptive Neighborhood (henceforth SBAN), a fixed size window is centered on each pixel of the reference image, but only selected pixels of this window are used to compute the correlation index. Any grey-level based correlation index can be modified using this technique. For example, the standard SAD expression is changed to :

$$C_l(x, y, s) = \sum_{i,j} \beta_{x,y}(i, j) \times |I_l(x + i, y + j) - I_r(x + s + i, y + j)| , \quad (1)$$

where the coefficient $\beta_{x,y}(i, j)$, is equal to one when the pixels of both images with window offsets i and j are used in the sum, to zero otherwise. This corresponds to defining a neighborhood with variable shape and size that can be adapted to the local image data.

In order to correspond to the same object as the center pixel $P_l(x, y)$, a pixel $P_l(x + i, y + j)$ is included in or excluded from the window according to a similarity criterion. If the two pixels are similar, $\beta_{x,y}(i, j)$ is set to one, otherwise it is set to zero. Many techniques can be used to define the similarity criterion, from simple ones like grey level comparison, to more complex ones like local texture analysis. In order to prove the efficiency of the SBAN approach



a) 3D scene

b) Left image

c) Right image

d) Fixed window

e) Adaptive window

Fig. 1. Fixed versus adaptive window

itself, we describe the algorithm for the simplest similarity criterion, i.e. the comparison of the grey levels. This is based on the assumption that a significant difference between the grey level means that the two pixels correspond to different objects of the scene.

In this case, $\beta_{x,y}(i, j)$ is set to one if the grey level $I_l(x+i, y+i)$ is close to the grey level $I_l(x, y)$ of the center pixel, more precisely if :

$$|I_l(x + i, y + j) - I_l(x, y)| \leq T_l(x, y) , \quad (2)$$

where $T_l(x, y)$ is the maximum acceptable difference between the grey levels. The value of this threshold must be related to the uniformity of the grey levels in the window. For example, it can be defined as :

$$T_l(x, y) = \frac{\sum_{i,j} |I_l(x, y) - I_l(x + i, y + j)|}{w \times w} . \quad (3)$$

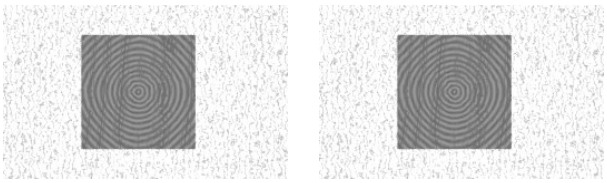
Like in standard methods, the disparity $d_l(x, y)$ is defined as the shift s giving the maximum (or minimum) value of $C_l(x, y, s)$. The disparity can be computed with subpixel accuracy by fitting a second degree curve to the correlation scores. In order to detect occlusions, the left-right consistency is used. For each pixel, if the disparity $d_l(x, y)$ computed using the left image as a reference is equal to the dis-

parity $d_r(x, y)$ computed using the right image as the reference, the solution is considered as correct. Otherwise the pixels are marked as occluded.

3. EXPERIMENTAL EVALUATIONS

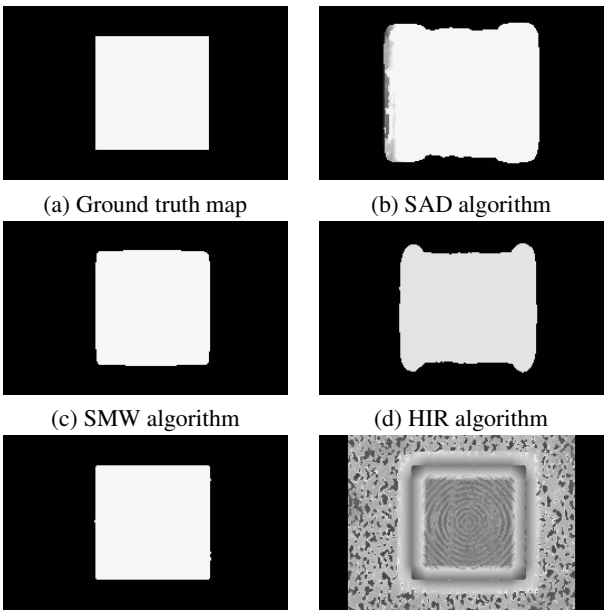
Two stereo pairs have been processed to demonstrate the effectiveness of our algorithm. We have compared our algorithm to the SAD on a fixed size window, to the SMW algorithm [3], and to the Hirschmüller (HIR) algorithm [4]. In each case, we present the disparity maps as grey level images, using a linearly stretched grey level range to improve readability : black for the minimum disparity, white for the maximum one.

For the first experiment, we have used the synthetic images shown in figures 2(a) and 2(b). The results are presented in figure 3. Two textured objects are present in the synthetic scene, that appear as a square and as the background in the images. Well defined edges correspond to depth discontinuities.



(a) Left image (b) Right image

Fig. 2. Synthetic stereo pair



(a) Ground truth map (b) SAD algorithm

(c) SMW algorithm (d) HIR algorithm

(e) SBAN algorithm (f) Pixels used

Fig. 3. Results for the synthetic pair (window size 27×27)

All the algorithms use a square window of size 27×27. In the areas corresponding to a single object, all the algorithms precisely estimate the disparity thanks to the large correlation window. The averaging effect, that introduces errors at depth discontinuities, is clearly visible in the disparity maps of the methods using a fixed size windows (cf. fig. 3(b), (c) and (d)). On the other hand, with the SBAN method (cf. fig. 3(e)), the estimated disparity map is very close to the ground truth, even near depth discontinuities. To better illustrate the behavior of the algorithm, we have included figure 3(f), which shows the number of pixels used in the sum defining the correlation index, a dark grey level indicating a small number of pixels. Few pixels are used near the borders of the square while many are retained in areas with an homogenous grey level, that are supposed to correspond to a single object.

Since the ground truth disparity map is available, a quantitative comparison is performed using percentages of pixels for which the disparity error is greater than one [7]. Two percentages are computed and presented in table 1, one for all non-occluded pixels, one for non-occluded pixels near depth discontinuities. This quantitative comparison shows that the error percentages increase with window size for SAD, SMW and HIR algorithms, but decrease with window size for the SBAN algorithm. With small windows, the four algorithms are almost equivalent since they yield high error percentages. These errors are caused by a lack of information in the correlation neighborhood. On the other hand, with large windows, the errors are mainly caused by mismatches near depth discontinuities, that are avoided with the SBAN method. This behavior is confirmed by the very low error percentages for the SBAN method for pixels near depth discontinuities.

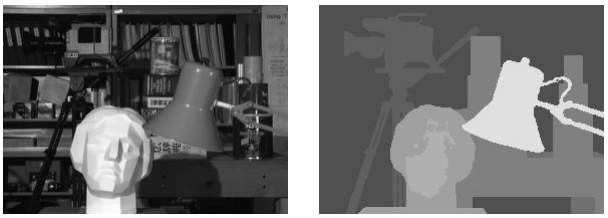
Algorithm	All pixels			Discontinuities		
	Window size			Window size		
	15	21	27	15	21	27
SAD	6.0	9.6	14.3	53.6	53.6	53.6
HIR	3.0	3.3	5.0	41.6	44.3	49.9
SMW	0.4	0.5	0.6	10.9	12.6	14.4
SBAN	0.3	0.3	0.3	9.1	7.8	7.3

Table 1. Errors for the synthetic pair (in %)

For the second experiment, we have used the Tsukuba stereo pair provided on Szeliski's[1] web page. Figure 4(a) shows the right image and figure 4(b) shows the ground truth map.

Figure 5 shows the disparity maps computed by the different algorithms. In this case, the visual comparison of the maps is more subjective. However, one can see that small details, like the lamp stem, or poorly contrasted objects, like

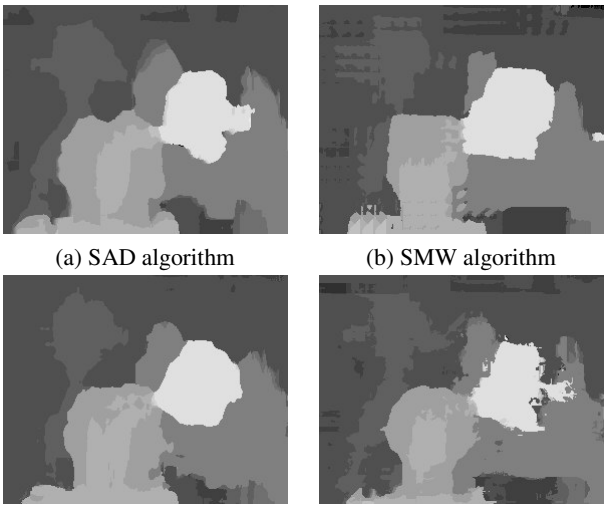
[1] http://research.microsoft.com/szeliski



(a) Right image (b) Ground truth map

Fig. 4. Tsukuba stereo pair

the camera in the background, are better processed by the SBAN method.



(a) SAD algorithm (b) SMW algorithm

(c) HIR algorithm (d) SBAN algorithm

Fig. 5. Results for the Tsukuba pair (window size 27×27)

The quantitative comparison presented in table 2 shows that SBAN algorithm have the smallest error percentages for both all non-occluded pixels and for pixels near depth discontinuities.

Algorithm	All pixels			Discontinuities		
	Window size			Window size		
	15	21	27	15	21	27
SAD	10.1	9.9	10.0	34.6	33.5	33.0
HIR	8.9	7.1	7.2	34.1	32.0	31.9
SMW	9.0	7.7	7.6	26.1	25.1	24.8
SBAN	7.1	6.9	6.7	19.0	18.8	18.5

Table 2. Errors for the Tsukuba pair (in %)

This second example shows the interest of the SBAN method, but also the necessity of defining an efficient similarity criterion. In this case, the result could become even better in homogeneous areas with a similarity criterion more precise than the basic difference between grey levels.

4. CONCLUSION AND OUTLOOKS

We have presented a new area-based algorithm for stereo matching using an adaptive window technique. The use of a large window allows for a stable computation of correspondences in texture-less areas. However, since the effective size and shape of the neighborhood are adjusted, blurring effects at discontinuities are avoided. This is a difference and an advantage with respect to the other classical area-based algorithms where the window size must remain small. Tests have shown the advantages offered by the SBAN algorithm, even with a trivial similarity criterion. We now work on the definition of other similarity criteria and study the implementation of the algorithm using specialized hardware for real-time processing.

5. REFERENCES

[1] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," in *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, Sacramento, CA, USA, Apr. 1991.

[2] C. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, July 2000.

[3] V. Roberto A. Fusiello and E. Trucco, "Symmetric stereo with multiple windowing," *International Journal of Pattern Recognition and Artificial Inteligence*, vol. 8, no. 14, pp. 1053–1066, dec 2000.

[4] H. Hirschmuller, "Improvements in real-time correlation-based stereo vision," in *Proceedings of IEEE workshop on Stereo and Multi-Baseline Vision*, Kauai, Hawaii, Dec. 2001.

[5] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proceedings of the European Conference of Computer Vision 94*, May 1994, pp. 151–158.

[6] Y. Zhang and C. Kambhamettu, "Stereo matching with segmentation-based cooperation," in *Proceedings of the Seventh European Conference on Computer Vision, ECCV'02*, Copenhagen, Denmark, June 2002.

[7] 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, pp. 7–42, apr 2002.