Towards correlation-based matching algorithms that are robust near occlusions

Sylvie Chambon Alain Crouzil
IRIT – Équipe TCI
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4, France
{chambon,crouzil}@irit.fr

Abstract

In the context of computer vision, matching can be done using correlation measures. This paper presents new algorithms that use two correlation measures: the Zero mean Normalised Cross-Correlation, ZNCC, and the Smooth Median Absolute Deviation, SMAD. While ZNCC is efficient in non-occluded areas and non-robust near occlusions, SMAD is non-efficient in non-occluded areas and robust near occlusions. The aim is to use the advantages of ZNCC and SMAD to deal with the problem of occlusions and to obtain dense disparity maps. The experimental results show that these algorithms are better than ZNCC-based algorithm and SMAD-based algorithm.

1. Introduction

One of the goals of stereovision is basically to find the third dimension from two images taken from two different angles. While solving this problem, two other subproblems occur: calibration and matching. Matching is an important task in computer vision because the accuracy of the 3D reconstruction depends on the accuracy of the matching. Matching is a difficult task because of: intensity distortions, noises, untextured areas, foreshortening and occlusions. A lot of algorithms have been proposed [16] to take these difficulties into account. The present paper only deals with matching using correlation measures [2].

In our previous work [5], the commonly used correlation measures have been presented and classified into five families; eighteen new correlation measures that are robust near occlusions have been proposed. In a scene, depth discontinuities induce occlusion problems. Pixels with different depth from the pixel being studied may be considered as outliers. Our robust measures are based on robust statistics tools because they are insensitive to outliers. We showed in [5] that SMAD is one of the most robust near occlusions whereas ZNCC is more efficient than SMAD in non-occluded areas. Consequently, we propose new algorithms that use both ZNCC and SMAD to obtain good results on the whole image. The most important difficulty of this kind of algorithm is to detect occlusions.

First, our definition of occlusions is given and some of occlusion detection methods are presented. Second, our algorithms are described. Third, we set up an evaluation protocol that compares all the algorithms. Finally, the results are discussed and conclusions drawn.

2. Occlusions: definitions and detection

The left and right images are denoted by I_v with $v = l, r$ and the following notations are used:

- The size of the correlation windows is $(2n+1)\times(2m+1)$ and $N = (2n+1)(2m+1)$, $n, m \in \mathbb{N}^*$;
- $\mathbf{p}_v^{i,j}$ with $v = l, r$ is the pixel in the image I_v at coordinates (i, j) and $I_v^{i,j}$ is the grey level of this pixel;

The following definitions are used (figures 1 and 2):

- **Occluded pixels** are pixels without correspondent:
$$\mathrm{O}(\mathbf{p}_v^{i,j}) = \begin{cases} 1 & \text{if } \mathbf{p}_v^{i,j} \text{ is an occluded pixel} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$
- **Occlusion area** contains all the occluded pixels in I_v:
$$\mathrm{OA}(I_v) = \left\{ \mathbf{p}_v^{i,j} \,|\, \mathrm{O}(\mathbf{p}_v^{i,j}) = 1 \right\} \quad (2)$$
- **Pixels near occluded pixels** are the pixels in the neighborhood of occluded pixels. This neighborhood is related to the size of the correlation window: it corresponds to the morphological dilation of the occlusion area using the correlation window as structuring element:
$$\mathrm{NO}(\mathbf{p}_v^{i,j}) = \begin{cases} 1 & \text{if } \mathrm{O}(\mathbf{p}_v^{i,j}) = 0 \text{ and } \mathrm{V}(\mathbf{p}_v^{i,j}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$
$$\text{where } \mathrm{V}(\mathbf{p}_v^{i,j}) = \sum_{p=-n}^{n} \sum_{q=-m}^{m} \mathrm{O}(\mathbf{p}_v^{i+p,j+q}) \quad (4)$$
- **Occlusion influence area** contains all the pixels near occluded pixels in I_v:
$$\mathrm{OIA}(I_v) = \left\{ \mathbf{p}_v^{i,j} \,|\, \mathrm{NO}(\mathbf{p}_v^{i,j}) = 1 \right\} \quad (5)$$
- **Whole occlusion area** is the union of occlusion area and occlusion influence area in I_v:
$$\mathrm{WOA}(I_v) = \left\{ \mathbf{p}_v^{i,j} \,|\, \mathrm{O}(\mathbf{p}_v^{i,j}) = 1 \text{ or } \mathrm{NO}(\mathbf{p}_v^{i,j}) = 1 \right\} \quad (6)$$

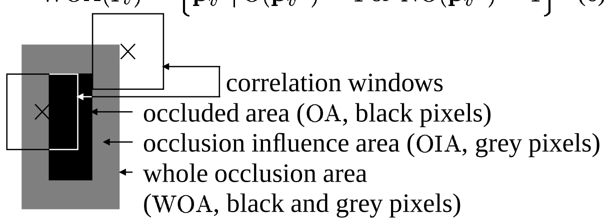


Figure 1. Occlusions: considered areas.

Dealing with occlusions is an important task in computer vision and a lot of methods have been proposed:

(a) (b) (c) (d)

Figure 2. (a)-(b) “Head and lamp” left and right images (c) Disparity map (d) Occlusions.

- Edge detection [18]: pixels that are near edge pixels can be considered as occluded pixels;
- Adaptive windows [10]: the shape of the correlation window depends on depth discontinuities detection;
- Dynamic programming with an occlusion constraint [3, 4, 8, 12, 13];
- Probabilistic approach with outlier modeling [9];
- Weighted correlation [11, 14]: an occluded pixel has a lower weight than a non-occluded pixel;
- Ordering constraint [6]: if $\mathbf{d}_v^{i,j}$ is the disparity vector of the pixel $\mathbf{p}_v^{i,j}$:

$$\begin{array}{l}\text{if } \mathbf{p}_l^{i_1,j_1} + \mathbf{d}_l^{i_1,j_1} = \mathbf{p}_r^{x_1,y_1} \text{ and } \mathbf{p}_l^{i_2,j_2} + \mathbf{d}_l^{i_2,j_2} = \mathbf{p}_r^{x_2,y_2} \\ \text{then } (i_1 - i_2)(x_1 - x_2) > 0 \text{ and } (j_1 - j_2)(y_1 - y_2) > 0\end{array} \quad (7)$$

- Uniqueness constraint [7]:

$$\begin{array}{r}\text{if } \mathbf{p}_l^{i_1,j_1} + \mathbf{d}_l^{i_1,j_1} = \mathbf{p}_r^{x,y} \\ \text{then } \forall i_2 \neq i_1, \forall j_2 \neq j_1,\ \mathbf{p}_l^{i_2,j_2} + \mathbf{d}_l^{i_2,j_2} \neq \mathbf{p}_r^{x,y}\end{array} \quad (8)$$

- Symmetry or bidirectional constraint [1, 6]:

$$\text{if } \mathbf{p}_l^{i,j} + \mathbf{d}_l^{i,j} = \mathbf{p}_r^{x,y} \text{ then } \mathbf{p}_r^{x,y} + \mathbf{d}_r^{x,y} = \mathbf{p}_l^{i,j} \quad (9)$$

3. Algorithms using two correlation measures

In the sequel, the following notations are used:

- Vectors $\mathbf{f}_v$ with $v = l, r$ contain the grey levels of the pixels in the left and right correlation windows: $\mathbf{f}_v = (\cdots I_v^{i+p,j+q} \cdots)^T = (\cdots f_v^k \cdots)^T$ where f_v^k is the element k of vector $\mathbf{f}_v$, with $p \in [-n; n]$, $q \in [-m; m]$, $k \in [0; N-1]$ and the transposed vector $\mathbf{f}$ is $\mathbf{f}^T$;
- L_P norms are noted: $\|\mathbf{f}_v\|_P = (\sum_{k=0}^{N-1} |f_k^v|^P)^{1/P}$ with $P \in \mathbb{N}^*$. The Euclidean norm is noted: $\|\mathbf{f}_v\| = \|\mathbf{f}_v\|_2$. The scalar product is defined by: $\mathbf{f}_l \cdot \mathbf{f}_r = \sum_{k=0}^{N-1} f_k^l f_k^r$ and the means are noted: $\overline{\mathbf{f}_v} = 1/N \sum_{k=0}^{N-1} f_k^v$;
- Ordered values of $\mathbf{f}$ are: $(f)_{0:N-1} \leq \cdots \leq (f)_{N-1:N-1}$.

The two correlation measures are the Zero mean Normalised Cross-Correlation, ZNCC [2], and the Smooth Median Absolute Deviation, SMAD [5], defined by:

$$\mathrm{ZNCC}(\mathbf{f}_l, \mathbf{f}_r) = \frac{(\mathbf{f}_l - \overline{\mathbf{f}_l}) \cdot (\mathbf{f}_r - \overline{\mathbf{f}_r})}{\|\mathbf{f}_l - \overline{\mathbf{f}_l}\| \|\mathbf{f}_r - \overline{\mathbf{f}_r}\|} \quad \text{and} \quad (10)$$

$$\mathrm{SMAD}(\mathbf{f}_l, \mathbf{f}_r) = \sum_{i=0}^{h-1} \left((\mathbf{f}_l - \mathbf{f}_r - \mathrm{med}(\mathbf{f}_l - \mathbf{f}_r))^2 \right)_{i:N-1} \quad (11)$$

The integer h [15] is the number of data taken into account[1].

Two groups of algorithms are defined. The common steps of the first group of algorithms are:

[1]To obtain a measure that tolerates 50% of outliers, we choose $h = \frac{N}{2}$.

1. To build the disparity maps (left-right and right-left) using ZNCC;
2. To locate occluded pixels with the symmetry constraint;

$$\mathrm{O}(\mathbf{p}_l^{i,j}) = \begin{cases} 1 & \text{if } \mathbf{p}_l^{i,j} + \mathbf{d}_l^{i,j} = \mathbf{p}_r^{x,y} \text{ and} \\ & \mathbf{p}_r^{x,y} + \mathbf{d}_r^{x,y} \neq \mathbf{p}_l^{i,j} \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

3. To locate the whole occlusion area;
4. For the whole occlusion area, to calculate new disparities using SMAD.

The problem is how to detect the whole occlusion area in the step 3. We propose the following possibilities:

- ALGO 1: to select the pixels detected as edge pixels[2];
- ALGO 2: to select the pixels detected by step 2;
- ALGO 3: to select the pixels in the conditional morphological dilation of the occlusion area with correlation window as structuring element. It is conditioned by the number of occluded pixels, with T a threshold[3]:

$$\mathrm{NO}(\mathbf{p}_l^{i,j}) = \begin{cases} 1 & \text{if } \mathrm{O}(\mathbf{p}_l^{i,j}) = 1 \text{ or } \mathrm{V}(\mathbf{p}_l^{i,j}) > T \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

- ALGO 4: to select the pixels like in ALGO 3 but the dilation is conditioned by the number of occluded pixels and edge pixels, with T_1 and T_2 two thresholds[4]:

$$\mathrm{NO}(\mathbf{p}_l^{i,j}) = \begin{cases} 1 & \text{if } \mathrm{O}(\mathbf{p}_l^{i,j}) = 1 \text{ or } (\mathrm{V}(\mathbf{p}_l^{i,j}) > T_1 \\ & \text{and } \mathrm{NE}(\mathbf{p}_l^{i,j}) > T_2) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

$$\text{where } \mathrm{NE}(\mathbf{p}_v^{i,j}) = \sum_{p=-n}^{n} \sum_{q=-m}^{m} \mathrm{E}(\mathbf{p}_v^{i+p,j+q}) \quad (15)$$

$$\text{with } \mathrm{E}(\mathbf{p}_v^{i,j}) = \begin{cases} 1 & \text{if } \mathbf{p}_v^{i,j} \text{ is an edge pixel} \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

The two steps of the second group of algorithms are to compute independently two disparity maps using ZNCC and SMAD and to merge these maps in order to obtain the final disparity map. The results obtained respectively using ZNCC and SMAD are noted by $_\mathrm{z}$ and $_\mathrm{s}$ subscripts. When the pixel is occluded, the disparity vector has negative values (noted $\mathbf{d}_\mathrm{n}$). In ALGO 5, these rules are used:

- if the measures obtain the same result, then it is kept;
- if one measure indicates the pixel is occluded then:

$$\mathbf{d}_l^{i,j} = \begin{cases} \mathbf{d}_\mathrm{n} & \text{if } \left(\mathrm{O}_\mathrm{z}(\mathbf{p}_l^{i,j}) = 1 \text{ and } \mathrm{V}_\mathrm{z}(\mathbf{p}_l^{i,j}) > \frac{N}{2}\right) \\ & \text{or } \left(\mathrm{O}_\mathrm{s}(\mathbf{p}_l^{i,j}) = 1 \text{ and } \mathrm{V}_\mathrm{s}(\mathbf{p}_l^{i,j}) > \frac{N}{2}\right) \\ \mathbf{d}_\mathrm{z}^{i,j} & \text{if } \left(\mathrm{O}_\mathrm{s}(\mathbf{p}_l^{i,j}) = 1 \text{ and } \mathrm{V}_\mathrm{s}(\mathbf{p}_l^{i,j}) \leq \frac{N}{2}\right) \\ \mathbf{d}_\mathrm{s}^{i,j} & \text{otherwise.} \end{cases} \quad (17)$$

- if the measures give two different disparities, then the disparity of the measure that obtains the most coherent disparities in the correlation window is kept:

$$\mathbf{d}_l^{i,j} = \begin{cases} \mathbf{d}_\mathrm{z}^{i,j} & \mathrm{V}_\mathrm{s}(\mathbf{p}_l^{i,j}) > \mathrm{V}_\mathrm{z}(\mathbf{p}_l^{i,j}) \\ \mathbf{d}_\mathrm{s}^{i,j} & \text{otherwise.} \end{cases} \quad (18)$$

[2]Here, step edge pixels are detected by the SDEF method [17].

[3]In our experimentation, $T = (3/10)N$.

[4]In our experimentation, $T_1 = (3/10)N$ and $T_2 = (5/100)N$.

4. Evaluation protocol

Eleven pairs of images with ground truth are used: a random-dot stereogram and ten real images proposed by Scharstein and Szeliski [16] and that can be found at: `http://www.middlebury.edu/stereo/data.html`. Seven of these images are made up piecewise of planar objects (posters, some with cut-out edges) and three images are complex scenes. Because of the lack of space, the results of only one pair (figure 2) are presented.

For the evaluation of the results, ten criteria are chosen:

- Percentage of correct and false matches (COR, FAL);
- Percentage of accepted matches (ACC): if the distance between the calculated and the true correspondent is one pixel then the calculated correspondent is accepted;
- Percentage of false positives and false negatives (FPO and FNE): the algorithm finds the pixel is matched whereas it is not matched and vice versa;
- Percentage of correct matched pixels in occluded areas: the results in the occlusion area (OA), the occlusion influence area (OIA) and the whole occlusion area are distinguished (figure 2);
- Execution time (T) and disparity maps: the clearer the pixel is, the closer the point to the image plane and the larger the disparity.

The three steps of the basic correlation algorithm used are, for each pixel in the left image:

1. The search area is determined in the right image;
2. For each pixel in the search area, the correlation score is evaluated;
3. The pixel giving the best score is selected.

For our algorithm, the size of the correlation window is 9×9 (the most suitable size for this kind of images found in [5]). The images are rectified so the search area is limited to the size 61×1 (30 pixels before and 30 pixels after the pixel of interest). Moreover, a symmetry constraint is added (equation 9). Pixels that do not respect the symmetry constraint are considered as occluded. Our algorithms are also compared with the algorithms proposed by Lan [14] and Kaneko [11]. These algorithms use reweighted correlation. Lan uses the Least Median of Squares to compute the weights and the Zero mean Sum of Squared Differences. Kaneko uses a non-parametric transform to compute the weights and ZNCC.

5. Experimental results

For all the images, with ALGO 2 to 5, the results of matching are improved, compared to the results obtained with the ZNCC-based algorithm, for the percentage of correct pixels, accepted pixels, false negatives and corrects pixels in WOA and OIA. If the results of ALGO 5 are compared to the results obtained with the SMAD-based algorithm, the percentage of correct pixels and corrects pixels in WOA and OIA are also improved. ALGO 2 gives better results than 3 and 4 for the percentage of correct and accepted pixels but gives poor results for the percentage of occluded pixels because ALGO 2 re-examines true occluded pixels and finds false positives whereas it does not re-examine any false positive. ALGO 4 is less efficient than ALGO 5 for all the criteria. The problem of ALGO 1 is that too much non-occluded pixels are detected as occluded and so the percentage of correct pixels is low (figure 3 and table 1). The Lan algorithm gives too much false negative and the Kaneko algorithm gives the same results as ZNCC. ALGO 5 is the most efficient but also the most expensive.

ALGO	COR (%)	ACC (%)	FAL (%)	FPO (%)	FNE (%)	WOA (%)	OA (%)	OIA (%)	T (s)
ZNCC	56.5	9.3	25.1	1.2	17.2	56.8	41.4	61.3	**15**
SMAD	44.8	**15.4**	32	**0.7**	22.5	71.8	67.1	73.2	160
Lan	41.2	10.4	**23.1**	0.5	35.2	67.6	**76**	65.2	850
Kaneko	56.3	9.3	25	1.3	17.5	57	42	61.4	115
1	55.2	9.4	24.8	1.1	18.9	57.1	49	59.4	28
2	58.6	10.8	28.8	1.4	**11.3**	66.2	35.7	75.1	65
3	57.8	9.9	26.1	1	15.2	67.5	55.2	71.1	82
4	58.1	10.5	27	1.1	14	66	48	71.2	93
5	**58.7**	11.2	28.2	1.1	12.2	**72.7**	48.5	**79.7**	181

Table 1. Results.

6. Conclusion

This work is particularly concerned with the occlusion problems. Our proposition is based on using two correlation measures, one classical, ZNCC, and one robust, SMAD. In ALGO 1 to 4, we considered the computation of an initial disparity map using ZNCC with the symmetry constraint and then we tried to locate the whole occlusion area and we applied SMAD only in this whole occlusion area. The best way to compute this whole occlusion area is to process a conditional morphological dilation of the occlusion area detected by the first matching (ALGO 3). ALGO 5 is based on the computation of the two dense disparity maps. This latter algorithm is the most efficient and the relative improvement is, for the percentage of correct pixels 2.2%, for false negatives 5%, and for correct pixels near occlusions 8.4%. However, this algorithm is the most expensive. Although ALGO 3 is less efficient than ALGO 5, it is less expensive so ALGO 3 is a good compromise.

One of the perspective of this work will be to decrease the execution time by using box-filtering or multiresolution [7]. The second perspective will be to improve the robustness by using adaptive windows [10].

References

[1] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez. Symmetrical dense optical flow estimation with occlusions detection. In *European Conference on Computer Vision*, volume 1, pages 721–735, Copenhagen, Danemark, May 2002.

[2] P. Aschwanden and W. Guggenbül. Experimental results from a comparative study on correlation type registration al-

gorithms. In Förstner and Ruwiedel, editors, *Robust computer vision: Quality of Vision Algorithms*, pages 268–282. Wichmann, Karlsruhe, Germany, March 1992.

[3] P. N. Belhumeur and D. Mumford. A bayesian treatment of the stereo correspondence problem using half-occluded regions. In *Computer Vision and Pattern Recognition*, pages 506–512, Champaign, Illinois, USA, June 1992.

[4] A. F. Bobick and S. S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, September 1999.

[5] S. Chambon and A. Crouzil. Dense matching using correlation: new measures that are robust near occlusions. In *British Machine Vision Conference*, volume 1, pages 143–152, Norwich, Great Britain, September 2003.

[6] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions : Empirical comparisons of four approaches. In *Computer Vision and Pattern Recognition*, volume 2, pages 466–473, Hilton Head Island, South Carolina, June 2000.

[7] O. Faugeras, B. Hotz, Z. Zhang, and P. Fua. Real time correlation-based stereo : Algorithm, implementation and applications. Research Report RR-2013, INRIA, August 1993.

[8] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. In *International Conference on Image Processing*, volume 14, pages 211–226, Washington, USA, October 1995.

[9] D. Hasler, L. Sbaiz, S. Süsstrunk, and M. Vetterli. Outlier modelling in image matching. *Pattern Analysis and Machine Intelligence*, 25(3):301–315, March 2003.

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

[11] S. Kaneko, Y. Satoh, and S. Igarashi. Using selective correlation coefficient for robust image registration. *Pattern Recognition*, 36(5):1165–1173, May 2003.

[12] S. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *Computer Vision and Pattern Recognition*, pages 103–110, Hawaii, December 2001.

[13] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, Vancouver, Canada, July 2001.

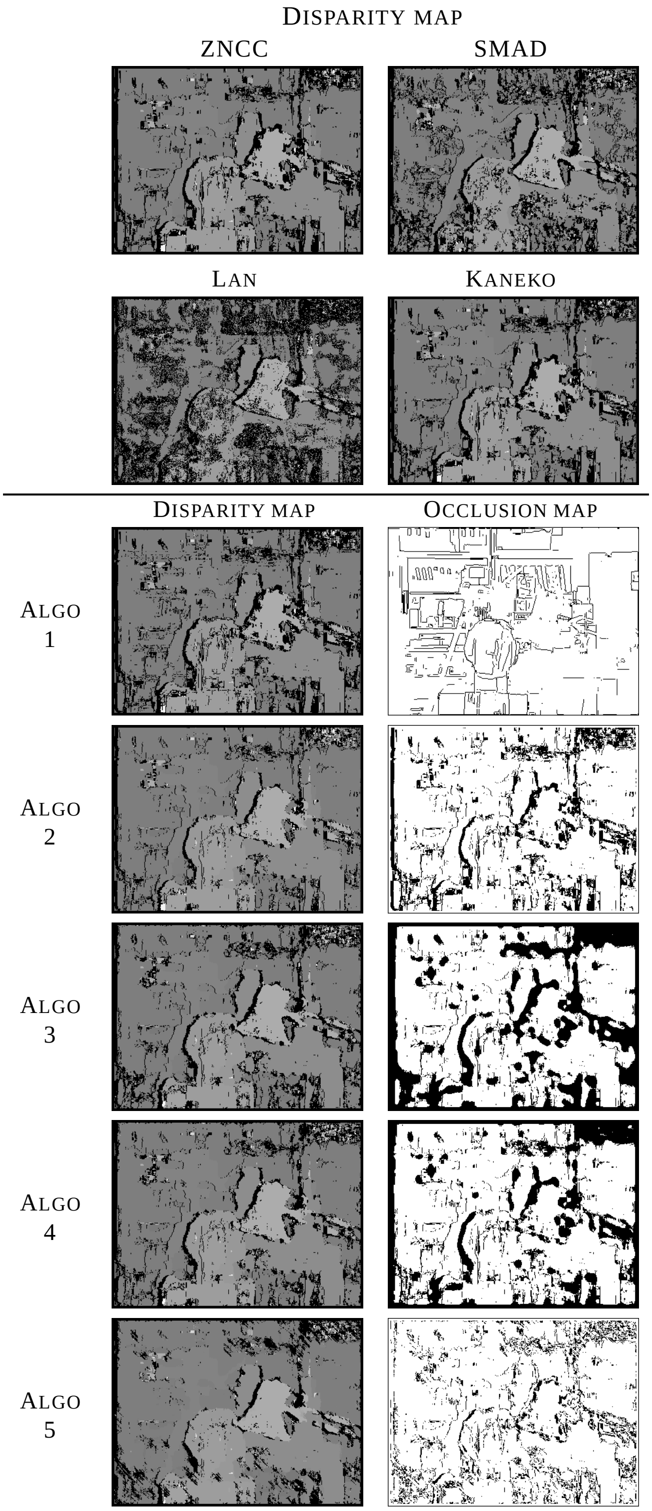
[14] Z. D. Lan and R. Mohr. Robust matching by partial correlation. In *British Machine Vision Conference*, pages 651–660, Birmingham, Great Britain, September 1995.

[15] P. J. Rousseeuw and C. Croux. *L_1-Statistical Analysis and Related Methods*, chapter Explicit Scale Estimators with High Breakdown Point, pages 77–92. Yadolah Dodge, Amsterdam, Holland, 1999.

[16] D. Scharstein and R. Szeliski. A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, April 2002.

[17] J. Shen and S. Castan. An optimal linear operator for step edge detection. *Computer Vision, Graphics, and Image Processing*, 54(2):112–133, March 1992.

[18] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *Pattern Analysis and Machine Intelligence*, 22(7):675–684, July 2000.



These images are available at:
`www.irit.fr/~Sylvie.Chambon/ICPR2004`.
For the occlusion maps, black pixels correspond to the pixels that are re-examined by SMAD.

Figure 3. Disparity and occlusion maps.