Robust matching by partial correlation *

Zhong-Dan LAN, Roger MOHR, Paolo REMAGNINO
LIFIA - INRIA Rhône-Alpes 46 avenue Félix Viallet
38031 Grenoble Cedex FRANCE
Zhong-Dan.LAN, Roger.MOHR, Paolo.REMAGNINO@imag.fr

Abstract

Stereo matching by correlation near occlusions is a very challenging problem. When a partial occlusion occurs, most of the standard methods fail to produce acceptable results. This is because the techniques used do not take into account the presence of the occluding region. We propose a robust technique which we call *partial correlation*. This technique makes use of the least median square method which has recently been used for vision problems such as robust surface reconstruction. It performs better than standard methods near occlusions. It works by first of all disambiguating between the occluding region and the object region in the template and in the candidate window. A binary weighted correlation is then performed on the object regions. We present a comparative study between our approach and two other techniques. Experiment results validate our approach.

1 Introduction

Area-based and feature-based matching techniques [7] [2] are the most commonly used techniques to determine correspondences between two images. In the literature, classical methods of area-based matching are:

- optical flow [8]
- Fourier transform [21]
- correlation for template matching [1] [4]

In this paper we will discuss only the last technique.

The use of correlation as a similarity measure between two signals is well-known. It is commonly used in stereo vision for the visual correspondence problem [16].

It is very hard to conceive a robust stereo template matching technique which can work near the occluding contours [16], [5] [10].

In the paper [10], Intille uses an energy functional which contains a term of similarity between corresponding pixels, a regularization term and some Ground

*This work has been performed within the research group MOVI which is common to CNRS, INRIA, INPG and UJF

BMVC 1995 doi:10.5244/C.9.65

Control Points (GCP) to cater for disparity discontinuity which occurs near occlusion.

Our opinion is that only the use of a *robust method* [9] [6] can overcome the matching problem caused by occlusions. We propose a template stereo matching method which provides good matches when other techniques fail. The key idea is to use a robust estimator to detect the portion of template which belongs to the object region and then perform a binary weighted correlation with the object region only.

The paper is organised as follows. Section 2 describes some related work. Section 3 describes the chosen approach. Section 4 illustrates some experiments and a final discussion concludes the paper.

2 Related work

Robust methods have been used in vision for quite some time in order to perform regression [13], clustering [11], or to compute the epipolar geometry [23]. Robust methods have also been used to improve matching techniques. For instance, Odobez and Bouthemy [18] use the *M-estimator* to compute the optical-flow, and Zabih [22] proposes a *non-parametric* transform to improve the correlation results.

It is very important to be able to find matches along the discontinuities, because discontinuities usually delimit the shape of an object. Unfortunately, traditional statistical similarity measures are not stable near discontinuities. This situation arises from the presence of multiple populations in a match window that includes a discontinuity.

We chose to describe in more detail Zabih's method and to make a comparison with our approach because it is a robust approach which works better than standard methods near occlusions.

Zabih proposes a new area-based approach for the visual correspondence problem. It is based on non-parametric local transforms followed by correlation. The use of transforms before correlation is similar to those proposed by Nishihara [17] and those of Seitz [1] [20]. Nishihara uses the sign bit of the image after convolution with a Laplacian while Seitz proposes the direction of the intensity gradient as the correlation primitive.

Let M be a square window of radius r and $N(P) = P \oplus M$ be a square of radius r centered at P, where $\oplus$ is the Minkowski summation operator. $N(P)$ will contain two distinct populations : the majority population, which arises from the object we are interested in and the minority population from the object across the discontinuity. Two distinctions can be made between the majority and the minority pixels:

- The majority pixels have a different disparity from the minority pixels
- The majority and minority pixels draw their intensities from different distributions

The presence of distinct populations within a match window is referred to as *factionalism*.

The fundamental idea behind Zabih's approach relies on local transforms based on non-parametric measures that are designed to tolerate the effect of *factionalism*. Non-parametric statistics [12] differ from the parametric technique in the way the ordering information among data is used rather than the data values themselves. Correspondence can be computed by transforming both images and then using correlation.

Zabih suggests two local non-parametric transforms. The first one, called *rank transform*, is a measure of local intensity; the second one, called *census transform*, is a summary of local spatial structure. The *rank* transform $R(P)$ is defined as the number of pixels in the local region whose intensity is less than the intensity of the center pixel. The *census* transform $R_\tau(P)$ is a mapping from the local neighborhood surrounding a pixel P to a bit string representing the set of neighboring pixels whose intensity is less than that of P.

Zabih's method is more stable near the disparity discontinuity than the classical correlation methods. A comparison of his methods and ours is provided in Section 4.

3 The partial correlation technique

Classical correlation techniques suffer from the problems of disparity discontinuity and highlights. When such problems occur, pixels in a region represent scene elements from two distinct intensity populations. We call this *partial occlusion*. Some of the pixels come from the object and some from other parts of the scene. We propose the *partial correlation* method to overcome this problem.

3.1 Constraints and likelihood used in our method

We propose a method based on the idea that a robust correlation measure can be computed in two steps. Firstly we identify the portion of the template which belongs to the object. The remaining portion represents the background. The selection is made using a robust method. In the second step, the object portion of the the template is matched against the second image. Section 3.2 describes in more detail the two steps.

In order to reduce the number of candidates, the *epipolar constraint* [16] and the *region of interest* constraint (ROI) [1] are used.

A likelihood function is required to measure the correlation between the template and the candidate windows. The classical methods and an experimental comparison between them is represented in [1]. We have chosen the zero mean sum of squared differences because of its relative simple form and its robustness to varying lighting conditions :

$$ZSSD(X, X+dX) = \sqrt{\frac{\sum_{\Delta \in W}((I_1(X+\Delta) - \bar{I}_1(X)) - (I_2(X+\Delta+dX) - \bar{I}_2(X+dX)))^2}{(2ulen+1)(2vlen+1)}}$$

where $X = (x, y)$, $\Delta = (u, v)$, $W = \{(u, v) | -ulen \leq u \leq ulen$ and $-vlen \leq v \leq vlen\}$, $dX = (dx, dy)$ means the disparity and $\bar{I}_i(X)$ are the mean values.

3.2 Our correlation model for partial occlusion

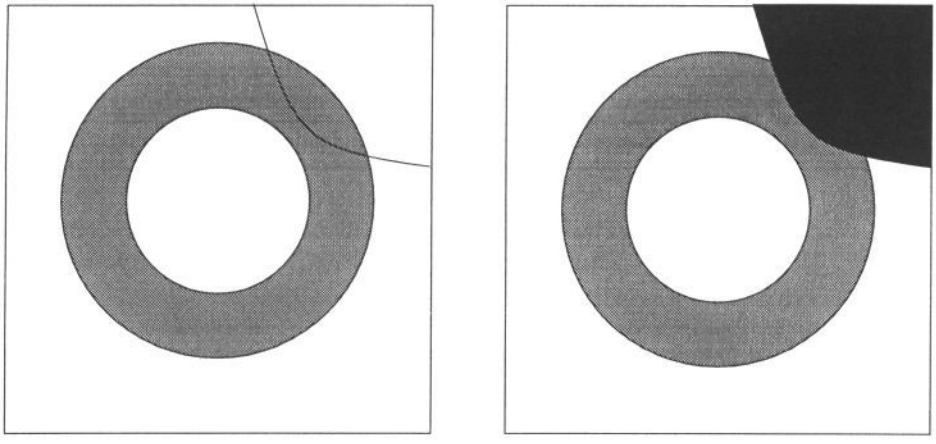

Figure 1: Template (left) and candidate window (right) under occlusion.

Consider the two windows (template I_1 and candidate I_2) in Figure 1, they match except for the top-right part of image I_1 and I_2. The goal of the first step of our method is to identify this occluded region. We assume that locally the image intensity signal undergoes an intensity shift from image I_1 to image I_2 (see Figure 2):

$$I_2(s,t) = I_1(s+dx, t+dy) + l + \epsilon \tag{1}$$

where (dx, dy) is the disparity of the point (x, y), and $\epsilon(s,t)$ is a Gaussian image noise.

Relation (1) does not hold for the occluded region.

If we consider that the center of the window of coordinates (x, y) is *not* occluded, we can add the following constraint :

$$I_2(x,y) \quad = \quad I_1(x+dx, y+dy) + l \tag{2}$$

We call such constraint the *center point constraint.* (CPC)

We use the least median square technique [19] to distinguish which pixels follow the model used. A Monte Carlo technique is used to compute the optimal model(if the CPC 2 is not used), that is the model which minimizes the median of the square residuals [19]. If the CPC 2 is used, no parameters remain to be estimated, so the model is got directly.

Such a model is consequently used to identify the inlier portion of the data. This is done by assigning a binary weight (*weight* in equation (3)) to each data point using the following formula :

$$weight(u,v) = \begin{cases} 1 & \frac{|r_{u,v}|}{\hat{\sigma}} \leq 2.5 \\ 0 & \frac{|r_{u,v}|}{\hat{\sigma}} > 2.5 \end{cases}$$

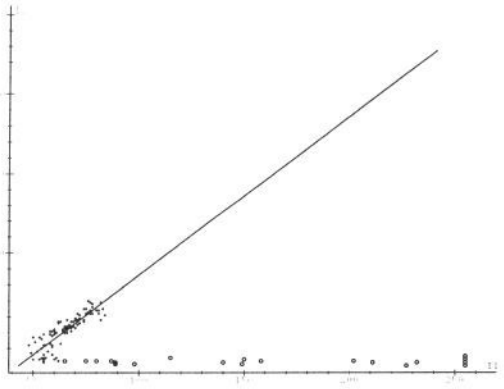

Figure 2: Relation between I_1 and I_2, the small circles represent the outliers

where $\hat{\sigma} = 1.4826(1 + \frac{5}{n-p})\sqrt{\underset{u,v}{\text{med}}\ r_{u,v}{}^2}$ is the estimation of the standard deviation of noise as described in [13]. Here n is the data number, p is the number of parameter and $r_{u,v}$ is the residual for the (u, v) data point. The factor 1.4826 is for consistent estimation in the presence of Gaussian noise and the term $\frac{5}{n-p}$ is recommended by Rousseeuw and Leroy [19] as a finite sample correction.

Having found the *occluded* portions, we perform the correlation on the remaining portions of the two windows. We call this technique *partial correlation.*

At this stage, a *weighted correlation technique* can be used to perform the *partial correlation. Partial correlation* is a special case of *weighted correlation*, as we leave $weight(u, v) = 1$ (inlier) or $weight(u, v) = 0$ (outlier). For example, the weighted $ZSSD$ can be expressed as follows :

$$ZSSD_w(X, dX) = \sqrt{\frac{\sum_{\Delta \in W}(((I_1(X+\Delta) - \bar{I}_1(X)) - (I_2(X+\Delta) - \bar{I}_2(X)))^2 weight(\Delta))}{\sum_{\Delta \in W} weight(\Delta)}} \quad (3)$$

where $\bar{I}_i(X) = \frac{\sum_{\Delta \in W} I_1(X+\Delta) weight(\Delta)}{\sum_{\Delta \in W} weight(\Delta)}$ for $i = 1, 2$

4 Experimental results and discussion

We present the efficiency of the partial correlation technique by comparing it with Zabih's and the classical one. In this paper, we show only our experiments using a planar scene occluded by an object. The use of a planar scene allows us to establish a one to one accurate mapping between the two images. We have tested the method using the translation relation (1) and the center point constraint (2). We call this method rzssdc. Only the translation relation has been used because

this method has been compared with $ZSSD$, which just takes into account a translation of the intensities.

We assume the pin-hole model for the camera. We use the epipolar constraint [16] and the ROI constraint ([1] in the computation of the correspondences and we use the homography between the two images to validate the result [16]. The following steps explain how the experiments were carried out :

- how to compute the epipolar geometry

  A simple setup is used. A calibration grid is placed in several positions at different depths from the two cameras with a relatively slanted orientation. Using a mode-based corner matching technique [3], the epipolar geometry between the two views is computed.

- ROI constraint

  We allow the disparity limits of ± 200pixels, while the size of the images is 512×512 (pixels).

- how to compute the homography

  The calibration grid is also used to compute the homography. An accurate approximation of the homography which maps the first view of a planar object onto the second is computed after having the correspondence of the grid corners. Such mapping is then used to verify the different correlation techniques. [3] [16]

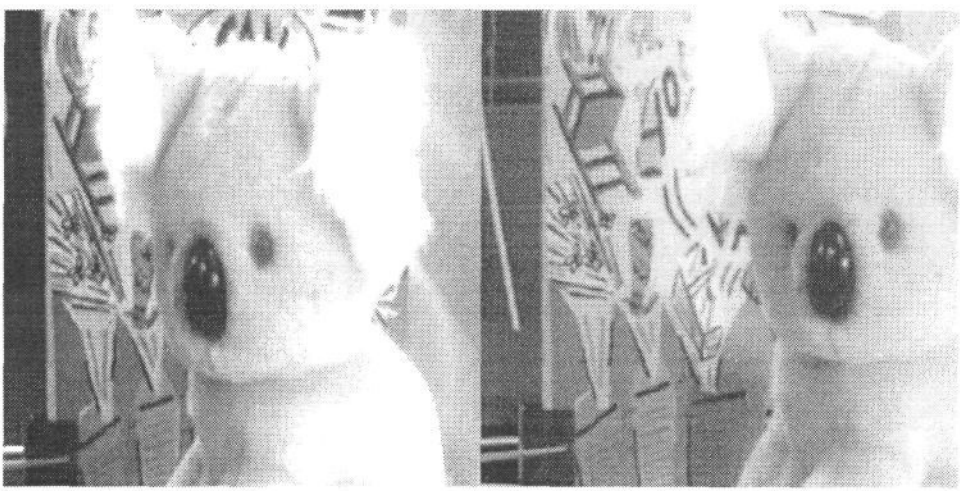

Figure 3: The two images to match

Many experiments were performed with this type of setup using different texture and occlusion. In this paper, only the experiments carried out with the two images of Figure 3 are presented. A planar scene was placed behind an object (the koala soft toy). Figure 4 shows two selected data sets on the left image in Figure

3 of the planar scene. The first one is made of points scattered along the occluding contour; the second one represents a rectangular region next to the occluding contour.

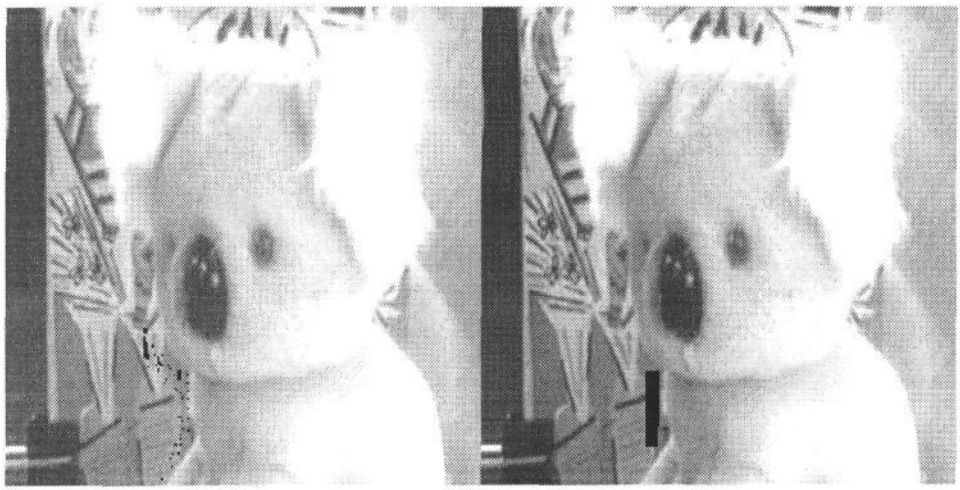

Figure 4: images and the points selected in it to compute the correspondence (black points). The results shown in Table 1 (resp. 2) have been obtained using the black points in the left (resp. right) image.

Our proposed method, Zabih's, and the classical method were run on these two data points. The candidates were constrained to move only along the epipolar line and inside the ROI. For each method, we compared the results found with the accurate value provided by the homography.

rank			
0–1	1–2	2–3	3–∞
77	14	0	49
census			
0–1	1–2	2–3	3–∞
89	12	0	39
rzssdc			
0–1	1–2	2–3	3–∞
118	10	0	12
zssd			
0–1	1–2	2–3	3–∞
56	0	0	84

Table 1: Results near occlusion

rank			
0–1	1–2	2–3	3–∞
129	15	0	122
census			
0–1	1–2	2–3	3–∞
185	6	0	75
rzssdc			
0–1	1–2	2–3	3–∞
234	29	2	1
zssd			
0–1	1–2	2–3	3–∞
71	1	0	194

Table 2: Results near occlusion

rank			
0–1	1–2	2–3	3–∞
11946	2512	299	968
census			
0–1	1–2	2–3	3–∞
12960	2237	105	423
rzssdc			
0–1	1–2	2–3	3–∞
11845	1985	193	1702
zssd			
0–1	1–2	2–3	3–∞
13140	1941	116	528

Table 3: Results far from occlusion

The results are reported in Table 1 and 2. For each method (i.e. : rank, census, rzssdc and zssd) the tables indicate the number of accurate matches, up to one

pixel error (good match); the number of matches within a distance between one and two pixels error (acceptable match); the number of matches at a distance of two to three pixels (failed matches) and the matches which lie more than three pixels away from their accurate position (false matches).

From these tables, we see that the rzssdc method gives the best result, the census is less good, the rank is even less so and the zssd is the worse.

However, Table 3 shows that if we choose some points far away from the occluding contour, the best results are obtained by using the census or the classical method.

We believe that our method uses a model which fitted to non contaminated but noisy data, may introduce false outliers. This is an acceptable explanation why the census and the classical methods work better far away from occlusions.

More work is needed to identify a more robust technique capable of performing well in all conditions.

5 Conclusion and future work

Occlusions carry important information in the visual correspondence problem, because they indicate where the disparity discontinuity occurs and they produce useful cues on the shape and location of the objects. However, an occlusion usually represents the limitation for most of the standard correlation methods.

In this paper, we discussed the occlusion problem and we proposed a new robust approach which we call the *partial correlation* method to overcome it. We compared our method with the standard correlation method and Zabih's correlation method. Experiments show that our method works better than other standard and non standard correlation methods when occlusions arise. However, when no occlusions are present, our method performs worse, even if it remains satisfactory.

For the moment, we think that the *partial correlation* represents only the beginning of the solution to the problem.

The fundamental limitation of this technique is the necessity for a tradeoff between the *similarity* and the *completeness* criteria : *Given a template, and several candidates, which one is the most similar to the template ?*

Perhaps there exists one candidate which is more similar to the template than others, which may be, on the other hand, more complete because they carry more information. The straightforward solution could be to take the one which is more similar when occlusions occur and the one which is more complete otherwise. This is an important issue not only in the problem of stereo matching by correlation, but also in the more general template matching problem.

Ongoing work has been targeted to automatically extract occluding contours in a given region using the presented technique. It works by performing the *partial correlation* in the selected region and by assigning a vote to the pixels on the border between the object and the background template portion. At the end of the process, each pixel of the selected region has a vote. The region can then be identified with a vote image which can be processed to extract the occluding contours. In practice, the vote image is closely related to a gradient image, therefore contours can be extracted by the technique of non-maximal suppression followed by double threshold.

The vote technique is strongly related to the notion of *consensus vision* [14] [15]. In the paper of Mintz [15], a vote technique is used to locate the edge. First results using this technique to locate occlusion contour have been obtained. They can be used to choose between the classical correlation and the partial correlation method we have developed. Moreover, the points on these contours can play the role of control points to build the disparity path introduced by Intille in his paper [10].

Acknowledgements

We would like to thank Pascal BRAND for his helpful input.

References

[1] P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In Förstner and Ruwiedel, editors, *Robust Computer Vision*, pages 268–282. Wichmann, 1992.

[2] N. Ayache. *Stereovision and sensor fusion.* MIT-Press, 1990.

[3] P. Brand and R. Mohr. Accuracy in image measure. In Sabry F. El-Hakim, editor, *Videometrics III*, volume 2350, pages 218–228, November 1994.

[4] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 1990.

[5] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 425–433. Springer Verlag, 1992.

[6] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics : the Approach Based on Influence Functions.* Wiley series in probability and mathematical. John Wiley and Sons, New York, 1986.

[7] R. Horaud and Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. IEEE *Transactions on PAMI*, 11(11):1168–1180, 1989.

[8] B. Horn and B. Schunk. Determining Optical Flow. Ai-memo 572, MIT, 1980.

[9] P.J. Huber. *Robust Statistics*, volume IX of *Wiley.* John Wiley, New York, 1981.

[10] S.S. Intille and A.F. Bobick. Disparity-space images and large occlusion stereo. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 179–186. Springer-Verlag, 1994.

[11] J.M. Jolion, P. Meer, and S. Bataouche. Robust clustering with applications in computer vision. *PAMI*, 13(8):791–802, August 1991.

[12] E. L. Lehman. *Nonparametrics: statistical methods based on ranks*. Holden-Day, 1975.

[13] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[14] P. Meer, D. Nintz, A. Montanvert, and A. Rosenfeld. Consensus vision. *AAAI-90*, July 1990.

[15] D. Mintz. Robust consensus based edge detection. *Computer Vision, Graphics and Image Processing: Image Understanding*, 59(2):137–153, March 1994.

[16] R. Mohr, P. Brand, and P. Remagnino. Correlation techniques in adaptive template matching with uncalibrated cameras. In *Vision Geometry III, SPIE's international symposium on photonic sensors & control for commercial applications*, volume 2356, pages 252–253, October 1994.

[17] H.K. Nishihara. PRISM, a pratical real-time imaging stereo matcher. Technical Report Technical Report A.I. Memo 780, Massachusetts Institute of Technology, 1984.

[18] J.M. Odobez and P. Bouthemy. Estimation robuste multi-échelle de modèles paramétrés de mouvement sur des scènes complexes. *International Journal of Computer Vision*, 11:419–430, 1994.

[19] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*, volume XIV of *Wiley*. J.Wiley and Sons, New York, 1987.

[20] P. Seitz. Using local orientational information as image primitive for robust object recognition. In *SPIE proceedings*, pages 1630–1639, 1989.

[21] J. Weng. Image matching using the windowed Fourier phase. *International Journal of Computer Vision*, 11(3):211–236, April 1993.

[22] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 151–158. Springer-Verlag, May 1994.

[23] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. Rapport de recherche 2273, INRIA, May 1994.