On Occluding Contour Artifacts in Stereo Vision

Radim Šára, Ruzena Bajcsy
University of Pennsylvania
GRASP Laboratory
3401 Walnut St., Philadelphia, PA, U.S.A.
radim@grip.cis.upenn.edu

Published in: *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, CVPR'97.* pp. 852-857. San Juan, Puerto Rico, June 1997.

Czech Technical University, Faculty of Electrical Engineering
Department of Control Engineering, Center for Machine Perception,
121 35 Prague 2, Karlovo náměstí 13, Czech Republic
FAX +420 2 24357385, phone +420 2 24357458, http://cmp.felk.cvut.cz

On Occluding Contour Artifacts in Stereo Vision

Radim Šára, Ruzena Bajcsy
University of Pennsylvania
GRASP Laboratory
3401 Walnut St., Philadelphia, PA, U.S.A.
radim@grip.cis.upenn.edu

Abstract

We study occluding contour artifacts in area-based stereo matching: they are false responses of the matching operator to the occlusion boundary and cause the objects extend beyond their true boundaries in disparity maps. Most of the matching methods suffer from these artifacts; the effect is so strong that it cannot be ignored. We show what gives rise to the artifacts and design a matching criterion that accommodates the presence of occlusions as opposed to methods that identify and remove the artifacts. This approach leads to the problem of measurement contamination studied in statistics. We show that such a problem is hard *given finite computational resources, unless more independent measurements directly related to occluding contours is available. What can be achieved is a substantial reduction of the artifacts, especially for large matching templates. Reduced artifacts allow for easier hierarchical matching and for easy fusion of reconstructions from different viewpoints into a coherent whole.*

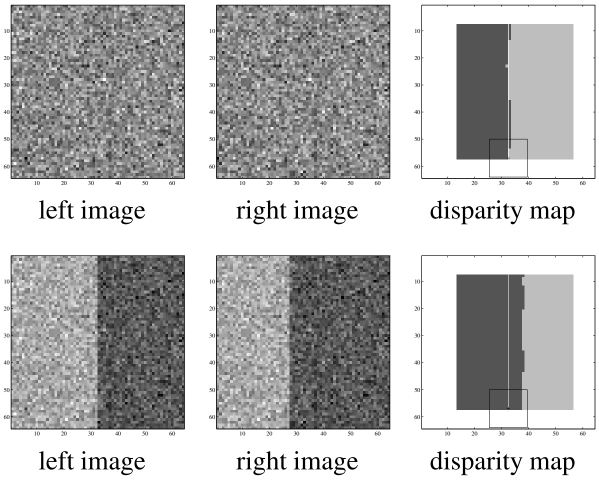


Figure 1. Two random-texture stereograms.

1. Introduction

Consider a binocularly viewed depth discontinuity, see the top row of Fig. 1. The corresponding disparity map recovered by area-based matching is shown far right. Both the surface textures are i.i.d. Gaussian noise of the same mean and variance and the pixels in correspondence are of the same value, so matching operator works quite well, exactly as expected in this 'textbook' example.

What happens when the average brightness of, say, the background texture changes? Interestingly, the computed disparity map is no longer accurate: *one surface extends beyond its true boundary,* see the bottom row of Fig. 1 (the white line marks the expected boundary in the disparity map, the matching window is outlined black). We used Sum of Squared Differences as the similarity measure. A more sophisticated criterion like Normalized Cross-Correlation (NCC) gives (almost) exactly the same results! The artifact is still there and its correlation value is as high as the value of any of the internal surface points!

In this paper we are going to investigate what causes the failure of the matching operator near the boundary and if it can be avoided, then how. We go beyond the standard problem of suppressing *occluded-area* artifacts.

1.1. Related work

A seemingly simple solution to the correct boundary reconstruction is to detect the boundary and then, in a *post-processing step,* rectify the false output of the matching process. The various methods may differ in how the boundary is detected (hypothesized).

The first group of methods use the intensity edge information in the input images. Cochran and Medioni [4] filter the disparity map by anisotropic diffusion that is stopped at intensity edges. They assume that the 'occlusion overhangs' will be smoothed out this way. Little and Gillett [12] locate discontinuities using information internal to the stereo module and then process this and additional information about boundaries in a post-processing step.

The second group of methods use no more information than is present in the recovered disparity map. Hoff and Ahuja [9] combine matching with surface interpolation and detect discontinuities and occluding contours as discontinuities in interpolated surfaces. Wildes [18] detects discontinuities in needle maps recovered from binocular disparity.

A different approach is to incorporate an occlusion process into the matching procedure. Belhumeur shows how to pose the problem as an optimization procedure [1]. Related approaches are reported, e.g., in [5, 13]. These methods try to modify the *decision procedure* rather than the similarity criterion which is used to compute the correlation values.

Bhat and Nayar [2] tried to modify the *similarity criterion.* They argue that robust matching method can generally improve the matching results by accommodating various unwanted transformations between the two images. They propose two modifications of Gideon and Hollister robust rank correlation coefficient as the matching operator.

Satoh and Ohta [17] increased the number of primary measurements: they used a 3×3 matrix of *color* cameras. They substantially reduce the proportion of 'unseen' regions. To obtain sharp boundaries, single color pixels are matched. The cameras are decomposed to eight subsets and the disparity is retained from that one giving the best match.

Okutomi and Kanade proposed a method that adaptively adjusts the matching window size based on image contents [15].

Various phenomena related to occluding contour are also studied in [3] and [14].

Once it is known where the boundary is, matching can be applied to both its 'sides' and the result with higher confidence retained. The difficulty is how to detect the boundary reliably. Even if it is done so, a curved boundary may still pose a problem, since it is displaced from its 'true' position due to the smoothing effect of the edge operator. To make the problem tractable, one has to *decide* where the boundary is and *later* reason about the matches. We want to merge these two stages by making no early decisions.

1.2. Our goal

We will not be locating occluded areas and boundaries; we want a procedure that *accommodates the fact* that they may occur in the observed world and *gives us correct answers* in their presence. (That also includes 'no answers' in the occluded areas.) This is close to the concept of outliers accommodation coined in robust statistics [8]. In this paper, we are going to explore this analogy.

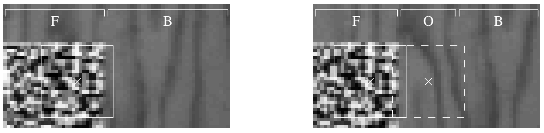
In the next section we answer the question of why the occlusion boundaries artifacts occur and in Section 3 we briefly describe methods that are suitable to solving the problem. Section 4 reports the results of their error evaluation. We summarize and give conclusions in Section 5. More details are given in the accompanying paper [16].

2. How do the artifacts develop?

Consider the portion of a real scene shown in Fig. 2. The scene consists of a corner of a randomly textured planar surface in front of a wooden background. Let the left-image matching window be at the position shown left in Fig. 2a (outlined) and let the right-image window be at the position marked by the solid outline shown right. It is easy to see that both windows are centered on the same point on the target. To evaluate the correlation between the windows, all the left-window pixel values are compared with the right-window values; the structure of their dependence is shown in the scatter plot in Fig. 3a. Every window contents comes from two different populations: the foreground and the background textures. The clusters correspond to foreground pixels plotted against the foreground pixels (F-F) and to background pixels plotted against the pixels in the occluded area (B-O). But only one of the clusters is relevant to the evaluation of the correlation value at the central pixel of the window! Generally, it is not known in advance which one should it be, since we do not know whether we observe an occlusion boundary, a texture boundary, or both.

Let now both the windows move *at the fixed disparity,* see Fig. 2b for a later stage (solid-outline windows). Some new clusters emerge (foreground–occluded area F-O, background–background B-B, see Fig. 3b), the fraction of the various pixels varies in the clusters, and so does the overall correlation value, see the solid curve in Fig. 4a. The correlation value should drop when the windows move away from the boundary but that does not happen in NCC. The correlation value stays high until the high-correlation cluster F-F disappears (see Fig. 2b).

The process should be symmetric if we tried to match the background (we are now comparing the contents of the window in the left image with the dashed-outline window in the right image in Fig. 2; note that both the windows are centered on the same *background* point in Fig. 2b). The correlation value should be low in early stages and it should rise later, when the fraction of correctly matched background pixels exceeds 50%. That does not happen until the uncorrelated F-O cluster disappears, see Fig. 3c, 3d and Fig. 4a, dashed curve. Matching background in Fig. 2b is complementary to matching foreground in Fig. 2a *but the clusters in Fig. 3c and 3d are not symmetric relative to those in Fig. 3b and 3a. The asymmetry—not the correlation absolute value—is responsible for the occlusion artifacts.* The profiles are sym-



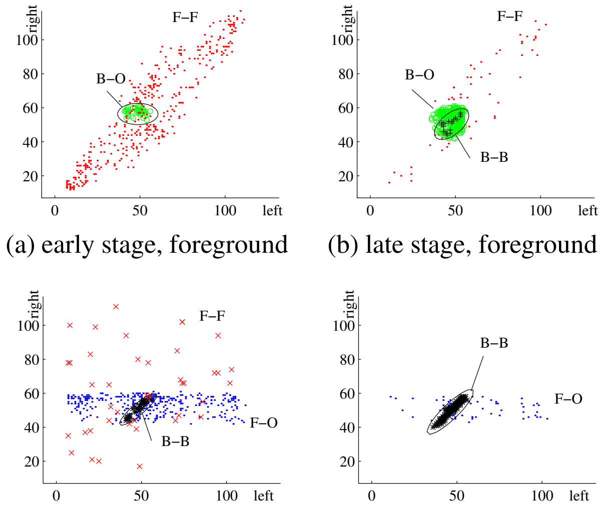
left image right image

(a) early stage, relative position -8 pixels

left image right image

(b) late stage, relative position $+8$ pixels

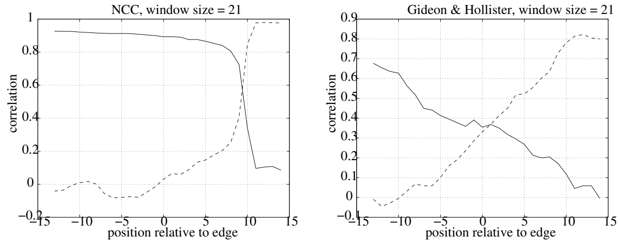
Figure 2. Matching near occlusion boundary.



(a) early stage, foreground (b) late stage, foreground

(c) early stage, background (d) late stage, background

Figure 3. Correlation structure.



(a) NCC (b) GHRRC

Figure 4. Correlation profile. The occlusion boundary is detected correctly if the two curves intersect at the relative position 0.

metric when the textures are the same, which explains why we got correct results in our first example from Fig. 1.

Note that the operator need not necessarily be robust in the sense that the correlation profile should drop/rise steeply near the boundary. Example of an operator that does not show this kind of robustness and still gives much better results than NCC is shown in Fig. 4b.

Besides the cluster determining the correlation, there are alien clusters whose influence must be eliminated. But since the various clusters are inseparable, their proper identification is not guaranteed. *We will consider a matching operator robust if it is able to accommodate the contamination,* i.e., give unbiased answers in its presence. The next section presents three of them and Section 4 compares their matching results.

3. Robust matching operators

Here we briefly review three correlation methods: one of them based on covariance matrix estimation and two of them based on rank correlations. Other methods are not analyzed in this paper although we experimented with them. They were the Minimum Volume Ellipsoid of Rousseeuw, the Huber's coefficient based on a robust variance estimator [10, p. 203], and the non-robust rank correlation methods (Spearman's ρ and Kendall's τ). All of them were clearly inferior to the methods presented here, based on the relative visual differences in the recovered disparity maps and on preliminary measurements of the accuracy on occlusion boundary. Complete evaluation was only performed for the selected methods.

3.1. Robust normalized cross correlation

The robust version of the NCC matching operator is based on robust covariance matrix estimation, for all details see [16].

For the initial robust location estimate $\mathbf{t}_0$, we use the LMedS estimator which has a breakdown point of 50%.

Given a family of p-dimensional elliptic distributions with probability density

$$f(\mathbf{x};\mathbf{t},\mathbf{V}) = |\det \mathbf{V}|\, f\left(\|\mathbf{V}(\mathbf{x}-\mathbf{t})\|\right), \qquad (1)$$

one wants to estimate the vector $\mathbf{t}$ and the pseudo-covariance matrix $\mathbf{S} = \mathbf{V}^T\mathbf{V}$ given n observations of $\mathbf{x}$ and the initial estimates of $\mathbf{t}_0$ and $\mathbf{S}_0$. The estimates are the solutions to the following set of equations:

$$\sum_{i=1}^{n} w(\|\mathbf{y}_i\|)\,\mathbf{y}_i = 0, \qquad (2)$$

$$\sum_{i=1}^{n} \left\{ u(\|\mathbf{y}_i\|)\frac{\mathbf{y}_i\mathbf{y}_i^T}{\|\mathbf{y}_i\|^2} - v(\|\mathbf{y}_i\|)\mathbf{E} \right\} = 0, \qquad (3)$$

where $\mathbf{y}_i = \mathbf{V}(\mathbf{x}_i - \mathbf{t})$; u, v, w are some scalar weighting functions, and $\mathbf{E}$ is a unit $p \times p$ matrix. We have chosen

$$\begin{aligned} w(r) &= \begin{cases} 1 & \text{if} \quad r \le s, \\ \frac{2\,s}{r} - 1 & \text{if} \quad s < r \le 2\,s, \\ 0 & \text{otherwise.} \end{cases} \\ u(r) &= r^2\, w^2(r), \\ v(r) &= \frac{1}{2}\, u(r). \end{aligned} \tag{4}$$

where $r = \|\mathbf{y}_i\|$. In our case, $p = 2$. The parameter s has to be chosen for a given class of images.

For the solution existence and uniqueness conditions, the reader is referred to [10, 8]. Unfortunately, redescending functions do not satisfy all of them.

The solution to (2)–(4) has to be found iteratively, see [10] for an algorithm.

3.2. Robust rank correlation

The components x_i and y_i of the n measurement vectors can be converted to their respective ranks $r(x_i)$ and $r(y_i)$. Ranks and rank correlations are invariant under any monotonic transformation of the measurements—a desirable property for image matching. For a very insightful overview of the standard methods see [11].

Gideon and Hollister [6] proposed a robust rank correlation coefficient, which we briefly describe here. It is assumed that there are no two equal values[1] among x_i and y_i. Let the vectors of ranks be simultaneously permuted so that $\pi(r(x_i)) = i$. Let $I(x) = 1$ if x is true and 0 otherwise. Then, for $i = 1, \ldots, n-1$, one defines

$$d_i = \sum_{j=1}^{i} I(i < \pi(r(x_{j2}))), \tag{5}$$

$$d_i^{\epsilon} = \sum_{j=1}^{i} I(i < n + 1 - \pi(r(x_{j2}))). \tag{6}$$

The Gideon and Hollister robust correlation coefficient is

$$\kappa_g = \frac{\max_i d_i^{\epsilon} - \max_i d_i}{\lfloor \frac{n}{2} \rfloor} \in [-1, +1], \qquad n > 1 \tag{7}$$

and it can assume $2\lfloor \frac{n}{2} \rfloor + 1$ different values. The value of κ_g can be computed in $O(n\, \log n)$ time [2]. Bhat and Nayar proposed a modification

$$\kappa_b = 1 - 2\frac{\max_i d_i}{\lfloor \frac{n}{2} \rfloor}, \tag{8}$$

which does not satisfy some desirable properties, for instance $\kappa_b(X, Y) \neq -\kappa_b(-X, Y)$, but is computationally less expensive.

[1] There are several methods how to treat these ties, see [7].

Rank correlations require ranking prior to computing the correlation. But the ranks can be substantially altered by the presence of contaminants in the data. Our experiments have shown that κ_g seems to be robust against permutations of ranks. To our knowledge, the breakdown point of κ_g has not been studied theoretically.

4. Experimental evaluation

The following errors were evaluated:

False positives are false disparity hypotheses in occluded areas where no disparity can be found. They are related to occluding contour and occluded area artifacts.

Total disparity error is the number of integer disparity map pixels that differ more that one level from ground-truth and have a positive correlation value. The total disparity error included mostly occluding contour artifacts.

Negatives are pixels where no disparity was found although the input data contained enough information to generate such hypothesis. The less the number of negatives the more efficient and unambiguous is the matching operator.

Note that the total disparity error is not the sum of the other two errors.

4.1. Experimental setup

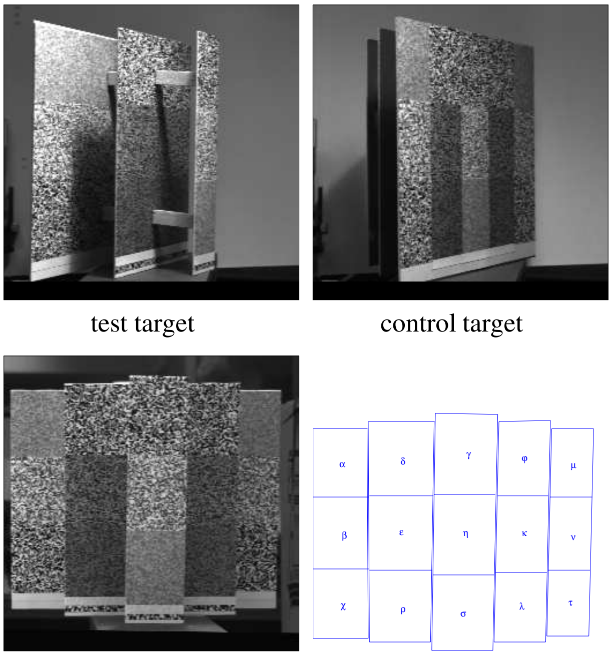
The test target is a staircase-like structure of known geometry, see Fig. 5. The target was designed so that it was easy to create a precise ground-truth disparity map semi-automatically (see [16]) and that all secondary phenomena affecting matching were eliminated (like projective distortions at excessively sloped surfaces; repetitive, low-contrast or missing texture; highlights; curved surfaces and boundaries; and shadowed areas).

The fifteen faces of the test target differ in the intra-face contrast of the random binary pattern. They are arranged so that all combinations of the relative inter-face contrasts are exhausted. We affixed the same texture pattern to a single planar surface and used it as a control target to verify the extend to which the matching operator is disturbed by the non-uniqueness of the artificial texture and by various texture boundaries in the absence of occlusions.

4.2. Error evaluation

The progression of occlusion artifacts with increasing size of a square matching window is shown in Fig. 6. The disparities were searched over the interval of $(-20, 15)$ in 256×256 images and all weak matches were eliminated using the disparity gradient limit and the ordering constraint.

The NCC-based matching was used as a control experiment. The artifacts develop rapidly with the increasing window size. There were no artifacts in the control target,



test target · control target

rectified left image · target structure

Figure 5. The test and the control targets.

only in the small-size matching windows a few matches were eliminated because of low correlation value.

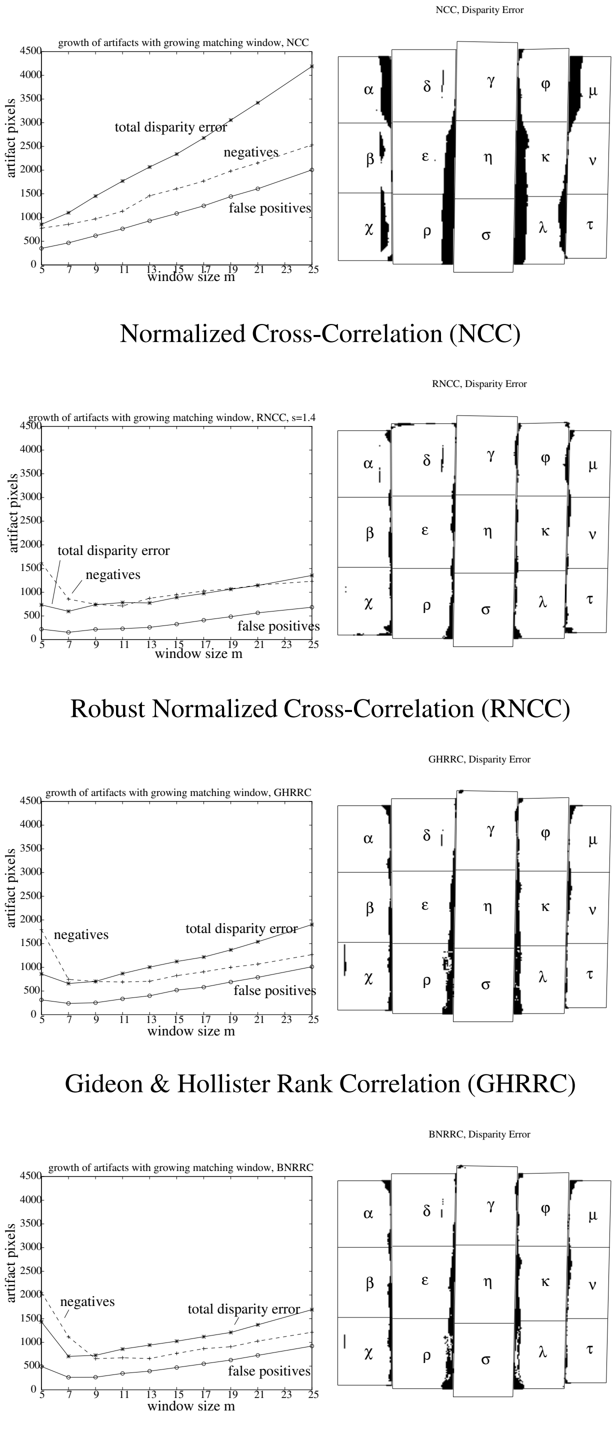
In RNCC[2], the artifacts develop too, but much slower. In small matching windows, the estimator is not efficient enough, so multiple matches and increased number of negatives (weak matches) in both the staircase and the control targets emerge. This is the reason why spurious local minima appeared in the plots.

Both robust rank matching operators perform slightly worse than RNCC. Of the two, GHRRC produces less artifacts and negatives.

The right column of Fig. 6 shows the detailed structure of the occlusion artifacts (black) for various matching operators and 25×25 matching window. The target structure is superimposed to show which artifacts are related to occlusion boundaries. A large matching window (not shown) was chosen to show how severe the problem could be. Notice that there are almost no artifacts at zero inter-face contrast, like in the $\delta - \gamma - \varphi$ transitions.

The fraction of false positives was $2.8\times$ smaller in RNCC and $1.4\times$ smaller in the rank-based operators than in NCC. The overall disparity error was $2.2\times$ smaller in RNCC and about $1.2\times$ smaller in the rank-based operators. Overall, RNCC was the best. The number of matching errors in the control plane was not significant.

[2]We used $s = 1.4$ and 3 iterations of the M-estimator.



Normalized Cross-Correlation (NCC)

Robust Normalized Cross-Correlation (RNCC)

Gideon & Hollister Rank Correlation (GHRRC)

Bhat & Nayar Rank Correlation (BNRRC)

Figure 6. Progression of occlusion artifacts (left). Total disparity error structure (right).

More error evaluation for different targets is described in [16]. Overall, the RNCC did 32% better than NCC and 13% better than the robust rank correlations in the simultaneous presence of smooth occlusions, specularities, shadows, and repetitive texture.

5. Summary and conclusions

The artifacts studied here have very high correlation value and are geometrically consistent with the surface they are adjacent to. They extend relatively far beyond the true boundary: up to a half of the matching window size. Once they are present in the disparity map, it is extremely hard to detect and remove them. The effect cannot be ignored, since 3-D reconstructions from multiple views based on wrongly recovered disparity maps become mutually inconsistent.

Since the occluding contour is viewed from two different viewpoints, the background in both images is different which introduces up to 50% of contaminant measurements in the data subset used for decision on correspondence. We found that the robust statistical methods we had tested were not able to cope with this sort of contamination and were therefore not suitable to solving the problem.

It is not just occluding contour artifacts but also other phenomena like the presence of surface specularities that lead to the contamination problem. By treating the problem as a general one, we hoped to reach robustness of the matching process *regardless* of the physical cause and without its explicit detection.

In the experiments, the various robust methods worked much better than the non-robust method. The price was high: the 280% increase in performance (for RNCC; in terms of false positives) is outweighed by the almost 100-fold increase in computational time!

Overall, the best results were obtained with RNCC. This method was least disturbed by the ambiguity caused by data distribution multi-modality falsely introduced by measurement contaminants.

We conclude that, although the occluding contour artifacts problem has been attacked at its root, we could not overcome the lack of information needed to resolve it. With finite computational resources one can only reduce the *expectancy* of the errors. What cannot be done without additional unambiguous information is to guarantee the non-existence of failure modes, even with infinite computational resources available. We want to put end to further research unless this additional information becomes at hand. The information could be of two kinds:

1. *Additional independent measurements* could be taken to help detect the correct position of the occluding contour.

2. *Additional constraints* could be employed to restrict the locations of candidate matches and to reduce the matching window size.

It would be very interesting to learn to what extent is the human visual system affected by the studied phenomenon.

Acknowledgments This work has been supported by the following grants: Army DAAH04-96-1-0007, DARPA N00014-92-J-1647, NSF SBR89-20230. First author has been partly supported by the Czech Grant Agency under the grant GAČR 102/95/1378.

References

[1] P. N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *Proc. 4th ICCV*, pp. 431–438, 1993.

[2] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. In *Proc. Computer Society CVPR*, pp. 351–357, 1996.

[3] R. Chung and R. Nevatia. Use of monocular groupings and occlusion analysis in a hierarchical stereo system. *Computer Vision and Image Understanding*, 62(3):245–268, 1995.

[4] S. D. Cochran and G. Medioni. 3-D surface description from binocular stereo. *IEEE Trans. PAMI*, 14(10):981–994, 1992.

[5] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. *Int. J. Computer Vision*, 14:211–226, 1995.

[6] R. A. Gideon and R. A. Hollister. A rank correlation coefficient resistant to outliers. *J. Am. Statist. Assoc.*, 82(398):656–666, 1987.

[7] J. Hájek and Z. Šidák. *Theory of Rank Tests*. Academic Press, New York, 1967.

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[9] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Trans. PAMI*, 11(2):121–136, 1989.

[10] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[11] W. H. Kruskal. Ordinal measures of association. *J. Am. Statist. Assoc.*, 53:814–861, 1958.

[12] J. J. Little and W. E. Gillett. Direct evidence for occlusion in stereo and motion. TR 90-5, Dept. of Computer Science, The University of British Columbia, Vancouver, 1990.

[13] A. Luo and H. Burkhardt. An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions. *Int. J. Computer Vision*, 15:171–188, 1995.

[14] J. Malik. On binocularly viewed occlusion junctions. In *Proc. ECCV*, pp. 167–174, 1996.

[15] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *Int. J. Computer Vision*, 7(2):143–162, 1992.

[16] R. Šára and R. Bajcsy. On occluding contour artifacts in stereo vision. GRASP Lab TR 411, University of Pennsylvania, Philadelphia, PA, 1997.

[17] K. Satoh and Y. Ohta. Occlusion detectable stereo using a camera matrix. In *Proc 2nd Asian Conf. on Computer Vision*, pp. 331–335, 1995.

[18] R. P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Trans. PAMI*, 13(8):761–774, 1991.