

Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches

Geoffrey Egnal and
Richard P. Wildes, *Member, IEEE*

Abstract—Binocular half-occlusion points are those that are visible in one of the two views provided by a binocular imaging system. Due to their importance in binocular matching as well as, subsequent interpretation tasks, a number of approaches have been developed for dealing with such points. In the current paper, we consider five methods that explicitly detect half-occlusions and report on a more uniform comparison than has previously been performed. Taking a disparity image and its associated match goodness image as input, we generate images that show the half-occluded points in the underlying scene. We quantitatively and qualitatively compare these methods under a variety of conditions.

Index Terms—Stereo matching, binocular half-occlusions, three-dimensional vision, empirical comparisons.

1 INTRODUCTION

OWING to the geometry of the situation, certain scene points will be visible in only one of the two views provided by a binocular imaging system. We refer to the image of these points as binocular half-occlusion points. They usually occur around object contours and other scene discontinuities. As such, these points have great potential to aid in image understanding tasks like distance estimation and object segmentation. It appears that the human visual system is capable of exploiting such information to great advantage [20], [13], [23], [1]. In contrast, computer vision has not been able to achieve similar success.

The basic geometry and perceptual significance of binocular half-occlusions has been known at least since the time of Leonardo DaVinci [25]. In contrast, the majority of computer vision algorithms for binocular stereo vision have ignored half-occlusions or treated them simply as noise in the matching process [3]. However, a number of approaches have emerged that deal with occlusions with various degrees of explicitness. Although the stereo matching techniques that deal with occlusions differ in many ways, it is possible to discern certain geometric assumptions and constraints that recur across these methods. In such a manner, it is possible to distinguish five major approaches:

1. Bimodality (BMD) is based on the principle that points around occlusion points will match both the occluded and occluding surface, creating a bimodal distribution in a local histogram of the disparity image. Previous research that has considered this approach includes [29], [21], [35].
2. Match Goodness Jumps (MGJ) assumes that badly matched points are occlusion points and that the matching algorithm itself can detect badly matched points. Previous research into this idea includes [28], [1].

• G. Egnal is with the GRASP Laboratory, University of Pennsylvania, 3401 Walnut St., Suite 300C, Philadelphia, PA 19104.
E-mail: gegrnal@gradient.cis.upenn.edu.

• R.P. Wildes is with the Department of Computer Science and Center for Vision Research, York University, 4700 Keele St., Toronto, Ontario M3J 1P3 Canada. E-mail: wildes@cs.yorku.ca.

Manuscript received 2 Mar. 2001; revised 17 Sept. 2001; accepted 9 Jan. 2002.
Recommended for acceptance by Z. Zhang.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113733.

3. Left-Right Checking (LRC) assumes that the left-based and right-based disparity images are viewing the same scene except at half-occlusions; therefore, values at corresponding points in the left and right disparity images should differ only in sign except at those points that arise from occlusions. Researchers that have examined this constraint include [34], [21], [8], [22], [16], [32].
4. Ordering (ORD) asserts that if two points match in a different order than they occur in the other view, then the point is an occlusion point. Previous research looking at this approach includes [2], [36], [24], [15], [12], [4].
5. The occlusion constraint (OCC) is based on the principle that a discontinuity in the left-based disparity image corresponds to an occlusion region in the right-based disparity image and vice versa. Previous research into this theory includes [15], [12].

Finally, we note that other approaches can be seen as dealing with occlusion not as part of image matching, but rather as a part of 3D reconstruction (e.g., [19], [31]), by assessing the degree to which the recovered 3D representation properly projects to the given images. We will not deal further with this last class of approaches in the present paper, as our emphasis is on the analysis of half-occlusion in conjunction with stereo matching.

In the light of previous work, our main contribution in the current paper is to report on an empirical evaluation of five methods of half-occlusion detection: BMD, MGJ, LRC, ORD, and OCC. The evaluation explicitly treats half-occlusions based on their underlying geometric derivation. Taking a disparity image and match goodness image as inputs, we generate images which show the half-occluded points in a scene and provide qualitative and quantitative evaluation. We employ laboratory images and real world images in our study, representing a small, but illustrative, subset of the infinity of possible scenes. We compare the occlusion methods under two different matching algorithms that generate the disparity and match goodness images (correlation-based scanline search (COR) and gradient-based plane-plus-parallax (PPP)). A preliminary version of this research has appeared previously [10].

2 ALGORITHM DEFINITIONS

We can subdivide the five occlusion detection methods into two groups. The first two algorithms (BMD and MGJ) detect occlusion borders, or where occlusion regions begin or end. The last three algorithms (LRC, ORD and OCC) label the entire half-occlusion region. Our comparisons accommodate for this difference; we only look for occlusion borders for the first group, and we look for the entire occlusion region for the second group.

2.1 Bimodalities in Disparity (BMD)

The idea behind BMD is that points in the horizontal disparity image that correspond to half-occlusion borders will have in their neighborhood disparity values that arise from both occluding and occluded surfaces. In such regions, the histogram of the disparity should be bimodal (see Fig. 1). We choose the peak ratio test to determine if there is a bimodality based on [29]. The peak ratio is the ratio of the second highest peak over the highest peak. If the peaks are of similar size, signifying bimodality, the peak ratio is closer to one. Analytically, one can express BMD as

$$\text{Bimodality} = \frac{\max(M_2)}{\max(M_1)},$$

where M_1 and M_2 are the first and second largest peaks in the local disparity field.

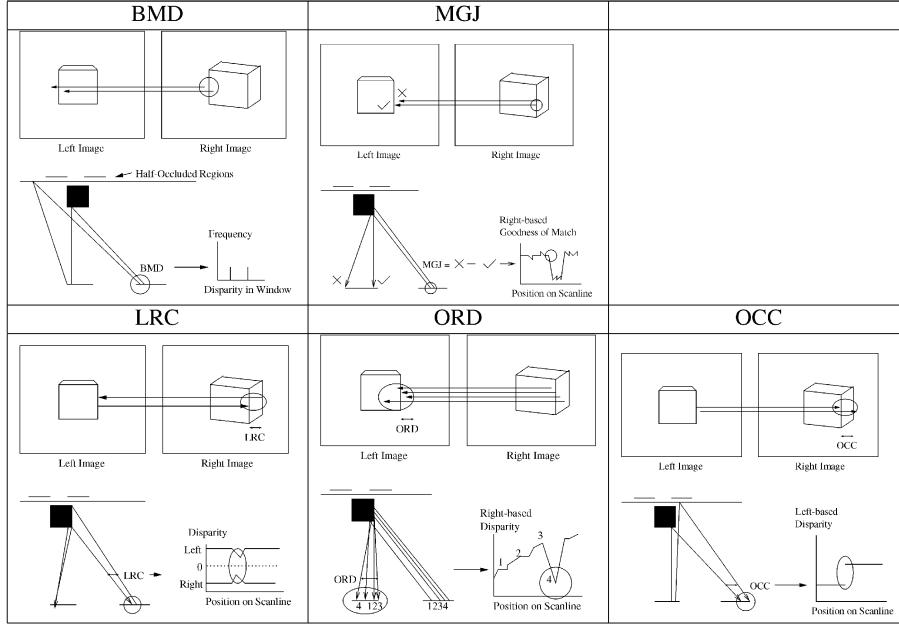


Fig. 1. Five Approaches to Detect Half-Occluded Points. The top row describes the border detection methods (BMD, MGJ), while the bottom row contains the region detection methods (LRC, ORD, and OCC). Each image shows three descriptions of how a given method would detect right occlusions. A stereo pair lies at top; an overhead view lies at bottom left, and a graph lies at bottom right that depicts either disparity along a scanline, disparity within a window, or match goodness along a scanline. An arrow indicates a match in a certain direction, and the checkmark/cross indicates high and low match score for the MGJ measure. All depicted matches lie along a single scanline.

2.2 Match Goodness Jumps (MGJ)

When a matcher is operating over data that arises from portions of a surface that are visible in both images of a binocular view, the goodness-of-match should be relatively high. In contrast, when the matcher considers data arising from a half-occlusion, the goodness-of-match should be relatively low since it is impossible to establish a correct correspondence. MGJ capitalizes on this state of affairs by detecting adjacent regions of high/low scores in goodness-of-match images (see Fig. 1). Analytically, one can express MGJ as

$$\text{Error} = \max(\overline{C}_x - \overline{C}_{x+w}, \overline{C}_x - \overline{C}_{x-w}),$$

where x is the horizontal coordinate of the current pixel, and \overline{C} is the summed match score within a window of size w .

2.3 Left/Right Checking Failures (LRC)

Since the left and right images are viewing roughly the same scene, the horizontal disparity images derived from matching right-to-left and left-to-right should be negatives of each other. The LRC hypothesis is that the points where the two images are not negatives of each other are occluded (see Fig. 1). More formally, if x_R matches $x'_L = x_R + d_{x_R}^R$, where x_R is a right coordinate and x'_L is an estimated left match at rightbased horizontal disparity $d_{x_R}^R$, then the right-based left/right error is,

$$\text{Error} = x_R - (x'_L + d_{x'_L}^L),$$

with $d_{x'_L}^L$ the left based horizontal disparity.

2.4 Ordering Constraint (ORD)

ORD states that if point A is to the left of point B in the left stereo image, then point A is also to the left of point B in the right image. In occluded regions, the false matches may be out of order and ORD can be used to label such points as occluded (see Fig. 1). Using the previous notation, we can express the ordering constraint as,

$$\text{Error} = \max(0, d_{x_R}^R - d_{x''_R}^R),$$

where x''_R represents the rightmost match so far.

2.5 Occlusion Constraint (OCC)

An ideal matcher produces a continuous disparity map between two continuous surfaces. However, near occluding boundaries, the disparity jumps from the occluding surface to the background, leaving occluded points in the opposite image unmatched. When the matching process jumps over points in the opposite image, OCC labels the unmatched points in the opposite image as occluded (see Fig. 1). Analytically, we can express the occlusion constraint near an occluding boundary as

$$\text{Error} = \max(0, d_{x_L+1}^L - d_{x_L}^L).$$

3 Empirical Results

For each stereo pair, we run two matching algorithms (COR and PPP, as described in the Appendix) and all five half-occlusion algorithms. The only adjustment to the algorithms between runs is the disparity search range of COR as detailed in the appendix. All other parameters are optimized by hand, but remain constant throughout the trials. The occlusion image results are plotted as grey-levels, with higher intensity corresponding to a stronger "occlusion signal," according to the algorithm definitions that were given in Section 2. We present only the results of right-based occlusion detection because the left-based results are similar and space limits such duplication.

The first data set is the University of Tsukuba [33] scene, which comes with a dense ground truth that is accurate within a pixel. The results of the five occlusion detection methods are shown in Fig. 2. As expected, the performance of the border detection (BMD and MGJ) and region detection algorithms (LRC, ORD, and OCC) is markedly distinct: the former signals the transitions into and out of occlusion regions while the latter signals the occluded region

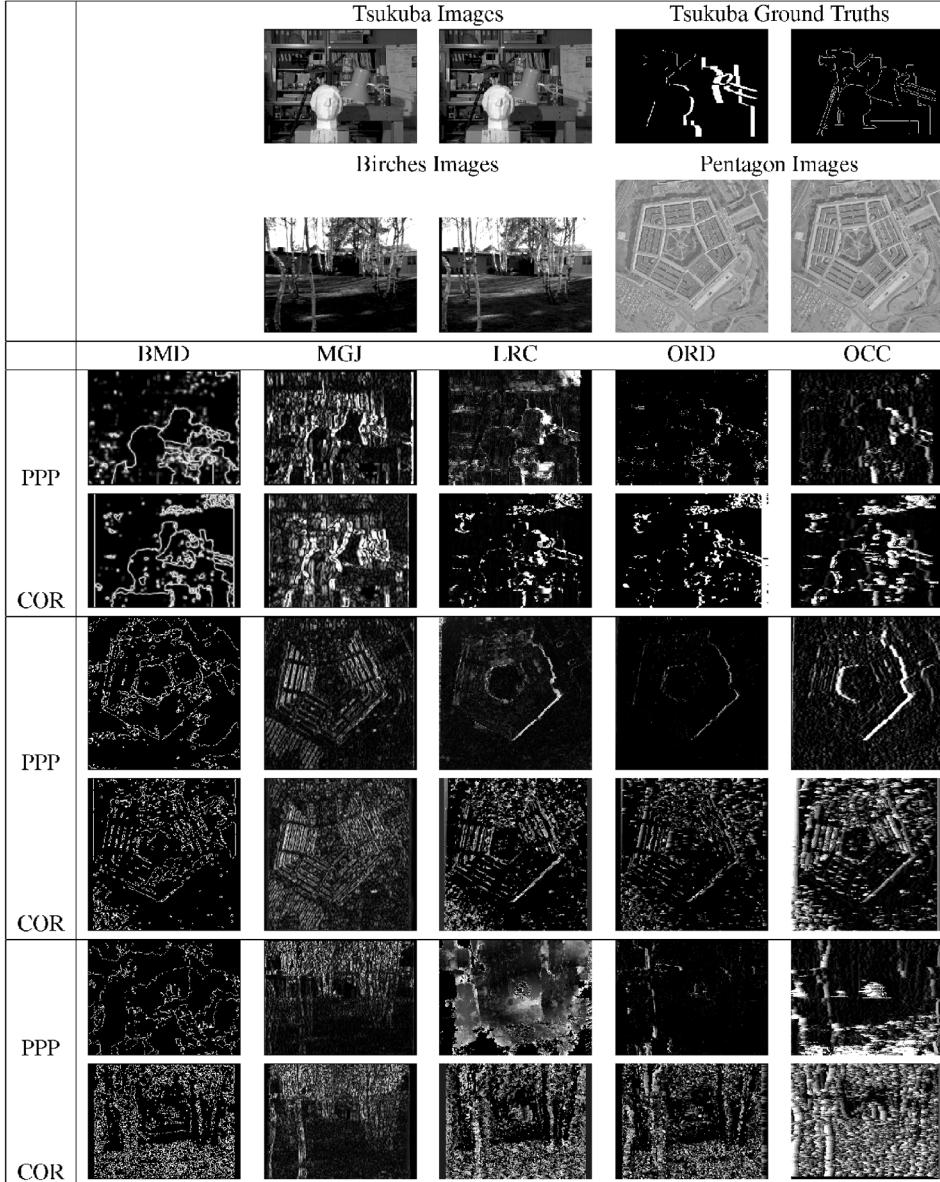


Fig. 2. Occlusion Borders and Regions Detected. The top two rows depict the original data sets. The results are in the following rows. From top to bottom, they are the Tsukuba data set, an aerial stereo pair of the Pentagon and a stereo pair depicting birch trees on a lawn. For each set, the 2×5 matrix of images shows the results of occlusion detection: Left-to-right—BMD, MGJ, LRC, ORD and OCC; top row—results using PPP matcher; bottom row—results using COR matcher. Brighter intensity values correspond to a stronger occlusion signal.

itself. Of the region detection algorithms, OCC detects the most occlusions, but falsely detects occlusions under COR on the left side of the head, where widely varying disparity estimates in the left-occluded region create spurious discontinuities. LRC comes close to OCC, but detects more false positives. While ORD is the most conservative measure, its accuracy is surprisingly low for two reasons. First, good matches can be labeled as occluded where noise-induced irregularities in matches appear: in essence, a bad match along a scanline can make ordering violations appear for some distance following the match. Second, ORD can miss those occlusion points that happen to match in order. Note that while the actual occlusion is at the right side of the objects, the different algorithms exhibit various degrees of ability to detect more general disparity discontinuities. The histogramming of BMD leads to detection of all occluding borders; the other methods are far less sensitive to nonoccluding borders. Interestingly, MGJ's false positives can differ from the others as MGJ is driven not by the

disparity per se, but rather by the goodness of match. For example, geometrically correct matches can still cause problems if there is significant photometric variation between matched features, as seen in the strong signals near the edges of the books on the bookshelf. Here, the edges have high match scores, while the interiors have low match scores, causing MGJ to pick up a false occlusion signal.

Quantitative analysis derives from the Tsukuba disparity ground truth, from which we have built a right-occlusion ground truth by labeling any unmatched pixel in the right image as a right-occlusion and disparity borders as an occlusion border (see Fig. 2). Using the ground truth, we compare the five methods (see Fig. 3). The graph plots hit rate and false positive rate at various thresholds. For parity, we limit the displayed false positive rate for all methods to 30 percent. After this point, BMD and PPP do not produce any more false positives before labeling all pixels occluded. Similarly, because there are only a certain number of

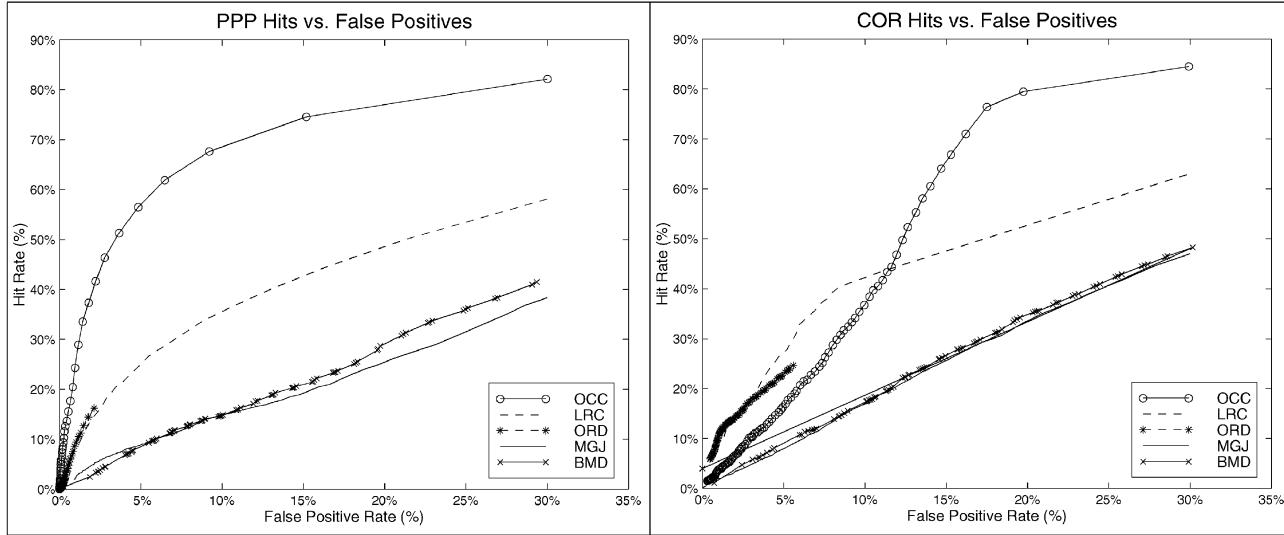


Fig. 3. Hit Rate versus False Positive Rates for the Tsukuba Data Set. The top graph shows the hit rate and false positive rate at various thresholds for the occlusion detection methods on the PPP matching results, while the bottom graph displays the methods' performance on the COR output. We limit the false positive rate for all methods to 30 percent, at which point, BMD and PPP will not produce any more false positives before labeling all pixels occluded. Similarly, because there are only a certain number of "out of order" matches, ORD can not label more pixels as false positives without labeling everything as occluded.

"out of order" matches, ORD can not label more pixels as false positives without labeling everything as occluded. For BMD and MGJ, we relax strict border detection and count hits within one pixel of the truth as correct to accommodate single-pixel mislocations. The most striking features of this comparison are as follows. OCC has the highest hit rate under PPP, but the second highest under COR. Presumably, the false matches on the left side of the head limit OCC's hit rate under COR. ORD is the most conservative measure, with the lowest overall false positive rate, but also the lowest hit rate of the region detection methods. BMD and MGJ perform similarly in detecting occlusion borders.

The last two tests use images of real world outdoor scenes. We consider the performance of the occlusion detection algorithms on two standard stereo vision data sets: the pentagon stereo pair from Carnegie Mellon University's VASC Image Database [7] and the birches stereo pair from the JISCT test set [5] (see Fig. 2).

For the Pentagon case, all of the algorithms perform reasonably in signalling the major occlusions arising from the rightmost edges of the building against the ground. Further, under PPP, the region detection algorithms, LRC, ORD, and OCC, provide markedly less spurious signal than the border detection algorithms. All of the algorithms deliver more false positives under COR. Presumably, this is due to its greater search range allowing it to find false matches, that subsequently are interpreted as occluded (see Fig. 4). There is considerable variability in the ability of the algorithms to detect the interior slits in the roof-top as occlusions. Under PPP, only MGJ shows any real ability to do this. This is due to the matcher producing uniform matches across the roof top, albeit with poor quality—a situation that only MGJ can capitalize on. Under COR, all algorithms are able to detect the occlusion-causing slits. Here, COR's larger search range works to its advantage in an interesting fashion: by looking further in an attempt to match the half-occluded regions, it produces matches that stand out geometrically to the BMD, LRC, ORD, and OCC approaches.

In the birches case, MGJ stands out as the best performer. Under both matchers, it produces a strong signal around the occluding borders of the trees while largely avoiding signalling occlusions in the other image regions. Under PPP, the remaining methods are only able to find correctly the scene's most pronounced occlusions. Further, additional signals are produced

in various erroneous regions (especially by BMD and LRC) as the matcher fails to establish reasonable correspondences between the stereo pair. Under COR, BMD, LRC, ORD, and OCC are much more sensitive in the correct occlusion areas. However, this sensitivity comes at the cost of greatly increased spurious signals. This overall pattern of results can be explained as follows: faced with the birches scene's considerable three-dimensional variation coupled with large areas of low contrast (e.g., the grassy areas), the matchers produce a large number of erroneous matches (in conjunction with more veridical matches on the trees). However, many of the false matches reproject well, especially for the case of COR. This combination of effects leads BMD, LRC, ORD, and OCC to mark the more curious (from a geometric standpoint) matches as occlusion effects; whereas, MGJ only signals occlusion where the

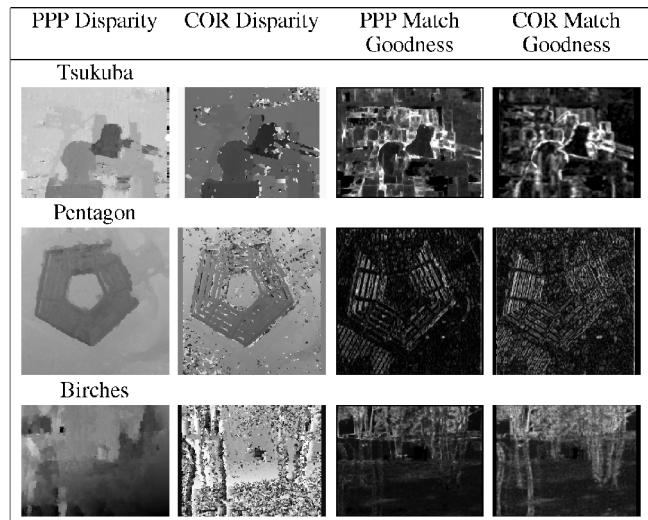


Fig. 4. Comparison of Disparity (left) and Goodness of Match Images (right) produced by two matching algorithms (PPP and COR). The images are of the Tsukuba scene, the Pentagon, and the Birches (top to bottom). COR is able to pick up greater disparities because of its wide search range. In contrast, PPP has a smaller search range. The match goodness images indicate that these two matchers have low match scores at similar points.

match is weak, which occurs as the matchers seek a correspondence for the half-occluded areas.

4 DISCUSSION

4.1 Algorithm Results

LRC consistently offers accurate occlusion labelling under many circumstances. It works best when the underlying scene geometry is frontoparallel because the matching algorithms work best under these situations. Although LRC is an expensive software solution, a parallel hardware implementation would make the method comparable in speed with the others considered in this paper. The method is slightly more sensitive to noise than the other algorithms because it relies on a matcher to perform correctly in two directions. If either match errs, the method will find an occlusion. However, the method has proven resistant to the noise levels presented in this paper.

ORD resists some of the matching errors that LRC labels as occlusions. In order to be labeled as an occlusion, the tight geometric assumption behind ORD demands that a bad match not only be bad, but be bad and out of order. ORD is straightforward to implement, requiring few parameters and little code. However, the algorithm has difficulties in the presence of extreme errors in the rightward direction and occluded matches that are in order. For these reasons, ORD might produce the fewest false positives when viewing surfaces whose noise has low magnitude. Although not seen in the examples presented, ORD will falsely label thin bars as occluded as they will appear out of order in correct matching, i.e., the double-nail illusion [17].

OCC labels occlusion regions even more accurately than LRC. Both quantitatively and qualitatively, OCC has demonstrated a higher hit rate than LRC. However, it often falsely labels occluded regions when operating in the occluded region of the opposite image. The severity of these false positives depends on which stereo matcher is in use. Under PPP, the false matches in occlusion areas are relatively smooth, and do not induce spurious OCC occlusions. Under COR, the matches in occluded areas are highly variable and discontinuous, which OCC falsely labels as occlusions in the opposite image. A two-way OCC check might eliminate such false positives, making OCC an extremely precise tool, albeit at the same computational expense as LRC.

BMD has the advantage that it detects all occluding edges, including discontinuities orthogonal to epipolar lines. It thereby serves as a more general indicator of 3D boundaries than the other methods. Although the method gives a good qualitative feel for the scene's occlusions, the method is highly sensitive to its parameters. The size of the window and the number of buckets in the histogram all have a large effect on the final appearance of the occlusion map. Given a certain type of scene, parameter tuning could increase the accuracy of BMD greatly.

Unlike the other methods, MGJ depends upon the internal assessment of the matching algorithm rather than the actual performance of the matching algorithm. The method uses this independence to distinguish some bad matches from occlusions. For instance, in areas of low spatial frequency, such as grass of the birch tree example, the matcher reports it has a good match, so MGJ detects no (false) occlusions, see Fig. 2. Other algorithms falsely label the bad match as an occlusion.

4.2 Overall Comparisons

The two matching algorithms have vastly different effects on the half occlusion methods. COR minimizes its sensitivity to noise by using large disparity search ranges. PPP makes use of gradient-based estimation (justified by assuming brightness constancy) and has a smaller search range, which can increase sensitivity to noise. As an example, all methods had more trouble under PPP than COR

with the large disparities found in the birches images. However, at single points, COR can be more prone to errors because of its larger search range, which allows it to incorrectly match similar points far away from the actual disparity. For example, under COR, OCC finds right occlusions where the left-based match in the left-occluded region has an incorrect and large disparity range. The most obvious effect of the different matching algorithms occurs when PPP cannot match under conditions that exceed its search range or is led astray due to violations of the brightness constancy constraint, thus negating any possible occlusion detection.

On an implementation level, the algorithms have many differences. ORD and OCC require the most straightforward implementation. The windowing used by the border detection methods is slightly cumbersome, and the need for two way calculations makes LRC and OCC the most expensive of the methods. However, a hardware implementation could perform the second matching process in parallel, making the methods more attractive. We also note that in our present study, half-occlusion detection was performed as post-processing on the stereo matchers' output; ultimately, it is of interest to more tightly couple the matching and occlusion detection processes.

The algorithms' performance depends on the imagery at hand. Simpler scenes (at an extreme, random dot stereograms) might have perfectly parallel epipolar lines and clean edges that drive the matchers to good results yielding correspondingly good results from the half-occlusion detectors. Another important issue is the overall amount of textural detail that is apparent in the images. In particular, image regions that are dominated by low spatial frequency detail, particularly if also of low contrast, cause many false positives in all the detectors, save MGJ. In contrast, better performance is generally achieved in regions with higher spatial frequency content.

In low frequency/low contrast areas, the matchers often produce geometrically incorrect matches, since locally the image features are weakly defined and small amounts of noise can favor a bad match, even though no occlusion is present. In such situations, BMD, LRC and ORD detect the curious match and label it (incorrectly) as due to an occlusion. In contrast, the goodness of match image in such low frequency areas often has consistently reasonable values as the match does, at least locally, appear to be of good quality. Therefore, MGJ passes over these regions without labeling them as indicative of half-occlusions. Finally, the distance between the foreground and background regions has an effect on occlusion detection, with larger separations generally yielding better performance, provided that the underlying matchers can capture the attendant disparity.

4.3 Summary

Overall, there does not appear to be a simple one-dimensional goodness ranking of the methods in the paper. At first glance at the quantitative data, OCC is the best algorithm. It has the highest hit rate and the lowest false positive rate within the test sets at hand. LRC appears a close second. However, the hit/false positive rates do not tell the whole story: OCC behaves badly in left-based occlusion regions when finding right-occlusions (and vice versa), while LRC does not perform well in areas of the scene that are dominated by low spatial frequency structure. Further, both of their implementations require a second matching process. Nevertheless, LRC does perform quite well in highly textured scenes, and OCC performs well given a matcher with smoother error characteristics. In contrast, in natural imagery, where low frequency spatial structure can be common, MGJ labels occlusions in a reasonable fashion even while being robust to the false positives that can plague the other methods in similar situations. For scenarios where 3D border detection is of primary interest, including those borders that are *not* manifest as half-occlusions, BMD performs well, albeit with a

tendency to over segment the scene. In contrast, ORD is overall the most conservative measure; still, it can produce false positives and is sensitive to the double-nail illusion.

In the final analysis, it may be desirable to have an integrated battery of half-occlusion detection mechanisms that span the scenarios of interest. For example, integrating LRC and OCC by using the left-based and right-based OCC would greatly decrease OCC's false positives. Further, if a priori knowledge or online diagnostics are available it could be advantageous to dynamically favor certain methods over others according to which are expected to yield the best performance in the situation at hand. Our results could serve to guide the design of such an integrated process.

APPENDIX

STEREO MATCHERS

A.1. Correlation Search-Based Matching (COR)

The first stereo matcher that we employ (COR) is a correlation-based scan-line matcher [3], [11]. We bring the epipolar lines into horizontal alignment by hand via a projective image warp. Following alignment, the images are filtered with a Laplacian operator [6], accentuating high frequency edges. Next, the matcher calculates 5×5 window size cross-correlation values for integral shifts, which are then smoothed. We use a search range of ± 17 pixels except for the birches image, which needed a search range of ± 27 . The disparity is the shift of each peak correlation value. Subpixel peak localization is made via interpolation with a quadratic polynomial. The match goodness image is defined as the correlation value at the final disparity.

A.2. Gradient-Based Plane Plus Parallax (PPP)

The second stereo matcher that we employ (PPP) is a gradient-based plane-plus-parallax method (as in [18], [26], [27].) The overall transformation between two images is parameterized as due to the displacement of a globally dominant plane, the epipole between the operative cameras and the parallax arising from points not lying on the plane. The disparity is the final parallax in the x and y direction. To begin, the algorithm uses a registration algorithm as a rough estimate of the epipole, allowing for an initial estimate of the planar parameters and parallax field. Next, the epipole is updated automatically based on the initial planar and parallax parameters. The estimation then iterates. To aid in the search for extended displacements, the estimation begins at a coarse level of a Laplacian pyramid and continues through finer levels of resolution [18]. We use five pyramid levels with three iterations per level. Similar to COR, the match goodness image is the square root of the sum of the squares of the differences in intensities between 5×5 windows brought into correspondence by the local parallax estimates.

A.3. Stereo Matcher Comparisons

In comparison to COR, PPP's use of gradient-based matching limits its search range. Embedding the matcher in a spatial coarse-to-fine approach effectively extends this search range, subject to the availability of spatial detail at coarse scales. In contrast, COR can search over larger ranges in full resolution, albeit at potentially great computational expense (see Fig. 4). Unfortunately, the greater search range provides the opportunity for COR to produce highly irregular matches, which could be ameliorated by enforcing continuity between scan-lines. The two matchers lead the occlusion detection algorithms to behave differently, as discussed in the main body of this paper. More general discussions of binocular stereo correspondence are available from a number of sources (e.g., [3], [9], [11] provide algorithmic overviews; [30] provides empirical

comparisons of several matchers; and [14] provides a psychophysical point of view).

ACKNOWLEDGMENTS

The authors would like to thank Yanlin Guo, Robert Mandelbaum and Harpreet Sawhney for their help. This study was begun at Sarnoff while G. Egnal was a summer intern and R.P. Wildes was a member of the technical staff.

REFERENCES

- [1] B. Anderson and K. Nakayama, "Toward a General Theory of Stereopsis: Binocular Matching, Occluding Contours, and Fusion," *Psychological Rev.*, vol. 101, no. 3, pp. 414-445, 1994.
- [2] H.H. Baker and T.O. Binford, "Depth from Intensity and Edge Based Stereo," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence*, pp. 631-636, 1981.
- [3] S.T. Barnard and M.A. Fischler, "Computational Stereo," *Computing Surveys*, vol. 14, no. 4, pp. 553-572, 1982.
- [4] P.N. Belhumeur, "A Bayesian Approach to Binocular Stereopsis," *Int'l J. Computer Vision*, vol. 19, no. 3, pp. 237-260, 1996.
- [5] R. Bolles, H. Baker, and M. Hannah, "The JISCT Stereo Evaluation," *Proc. Defense Advanced Research Projects Agency (DARPA) Image Understanding Workshop*, pp. 263-274, 1993.
- [6] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," *Trans. Computers*, vol. 31, pp. 532-540, 1983.
- [7] R. Bolles, H. Baker, and M. Hannah, "The JISCT Stereo Evaluation," *Proc. Defense Advanced Research Projects Agency (DARPA) Image Understanding Workshop*, pp. 263-274, 1993.
- [8] C. Chang, S. Chatterjee, and P.R. Kube, "On an Analysis of Static Occlusion in Stereo Vision," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 722-723, 1991.
- [9] U.R. Dhond and J.K. Aggarwal, "Structure From Stereo—A Review," *IEEE Trans. Systems, Man, and Cybernetics* vol. 19, no. 6, pp. 1489-1510, 1989.
- [10] G. Egnal and R.P. Wildes, "Detecting Binocular Half-Occlusions: Empirical Comparisons of Four Approaches," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. II, pp. 466-473, 2000.
- [11] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, Cambridge, Mass.: MIT Press, 1993.
- [12] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusion and Binocular Stereo," *Int'l J. Computer Vision*, vol. 14, pp. 211-226, 1995.
- [13] B. Gillam and E. Borsting, "The Role of Monocular Regions in Stereo Displays," *Perception*, vol. 17, pp. 603-608, 1988.
- [14] I.P. Howard and B.J. Rogers, *Binocular Vision and Stereopsis*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [15] S.S. Intille and A.F. Bobick, "Disparity-Space Images and Large Occlusion Stereo," *Proc. European Conf. Computer Vision*, pp. 179-186, 1994.
- [16] K. Konolige, "Small Vision Systems: Hardware and Implementation," *Proc. Int'l Symp. Robotics Research*, 1997.
- [17] J.D. Krol and W.A. van der Grind, "The Double-Nail Illusion," *Perception*, vol. 11, pp. 615-619, 1982.
- [18] R. Kumar, P. Anandan, and K. Hanna, "Direct Recovery of Shape from Multiple Views: A Parallax Based Approach," *Proc. 12th IAPR Int'l Conf. Pattern Recognition*, pp. 685-688, 1994.
- [19] K.N. Kutulakos and S.M. Seitz, "A Theory of Shape by Space Carving," *Int'l J. Computer Vision*, vol. 38, no. 3, pp. 199-218, 2000.
- [20] R.B. Lawson and W.L. Gulick, "Stereopsis and Anomalous Contour," *Vision Research*, vol. 7, pp. 271-297, 1967.
- [21] J. Little and W. Gillett, "Direct Evidence for Occlusion in Stereo and Motion," *Proc. European Conf. Computer Vision*, pp. 336-340, 1990.
- [22] A. Luo and H. Burkhardt, "An Intensity-Based Cooperative Bidirectional Stereo Matching with Simultaneous Detection of Discontinuities and Occlusions," *Int'l J. Computer Vision*, vol. 15, pp. 171-188, 1995.
- [23] K. Nakayama and S. Shimojo, "DaVinci Stereopsis: Depth and Subjective Occluding Contours from Unpaired Image Points," *Vision Research*, vol. 30, pp. 1811-1825, 1990.
- [24] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139-154, 1985.
- [25] *Selections from the Notebooks of Leonardo da Vinci*. J. Richter, ed. Oxford U.K.: Oxford Univ. Press, 1977.
- [26] H.S. Sawhney, "3D Geometry from Planar Parallax," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 929-934, 1994.
- [27] A. Shashua and N. Navab, "Relative Affine Structure," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 483-489, 1994.
- [28] T. Smitley and R. Bajcsy, "Stereo Processing of Aerial Urban Images," *Proc. Int'l Joint Conf. Pattern Recognition*, pp. 405-409, 1984.
- [29] A. Spoerri and S. Ullman, "The Early Detection of Motion Boundaries," *Proc. Int'l Conf. Computer Vision*, pp. 209-218, 1987.
- [30] R. Szeliski and R. Zabih, "An Experimental Comparison of Stereo Algorithms," *Proc. Int'l Workshop Vision Algorithms*, pp. 1-19, 1999.
- [31] H. Tao, H.S. Sawhney, and R. Kumar, "A Global Matching Framework for Stereo Computation," *Proc. Eighth Int'l Conf. Computer Vision*, pp. 532-539, 2001.

- [32] R. Trapp, S. Drue, and G. Hartmann, "Stereo Matching with Implicit Detection of Occlusions," *Proc. European Conf. Computer Vision*, pp. 17-33, 1998.
- [33] Univ. of Tsukuba Stereo Pair, obtained from Richard Szeliski at <http://www.research.microsoft.com/szeliski/stereo/> Website current as of Sept. 2001.
- [34] J. Weng, N. Ahuja, and T.S. Huang, "Two-View Matching," *Proc. Int'l Conf. Computer Vision*, pp. 64-73, 1988.
- [35] R.P. Wildes, "Direct Recovery of Three-Dimensional Scene Geometry from Binocular Stereo Disparity," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 761-774, 1991.
- [36] A.L. Yuille and T. Poggio, "A Generalized Ordering Constraint for Stereo Correspondence," Artificial Intelligence Lab 777, MIT, Cambridge, Mass. 1984.

▷ For more information on this or any other computing topic, please visit our Digital Library at <http://computer.org/publications/dlib>.

Bidirectional Deformable Matching with Application to Handwritten Character Extraction

Kwok-Wai Cheung, *Member*,

IEEE Computer Society,

Dit-Yan Yeung, *Member*, *IEEE*, and

Roland T. Chin, *Member*,

IEEE Computer Society

Abstract—To achieve integrated segmentation and recognition in complex scenes, the model-based approach has widely been accepted as a promising paradigm. However, the performance is still far from satisfactory when the target object is highly deformed and the level of outlier contamination is high. In this paper, we first describe two Bayesian frameworks, one for classifying input patterns and another for detecting target patterns in complex scenes using deformable models. Then, we show that the two frameworks are similar to the forward-reverse setting of Hausdorff matching and that their matching and discriminating properties are complementary to each other. By properly combining the two frameworks, we propose a new matching scheme called bidirectional matching. This combined approach inherits the advantages of the two Bayesian frameworks. In particular, we have obtained encouraging empirical results on shape-based pattern extraction, using a subset of the CEDAR handwriting database containing handwritten words of highly varying shape.

Index Terms—Model-based segmentation, deformable models, Bayesian inference, bidirectional matching, Hausdorff matching.

1 INTRODUCTION

To achieve integrated segmentation and recognition in complex scenes, the model-based approach has widely been accepted as a promising paradigm. For example, one can search for the presence of a rigid object in an input image by optimizing some data mismatch measure with respect to the geometric transformation applied to the model. However, if the object of interest is *nonrigid*, the potential shape variations can no longer be described by a compact set of transformation parameters. Instead, more flexible representations, commonly called *deformable models*, are required. Extracting nonrigid shapes using deformable models is known to be highly ill-posed. Very often, regularization techniques are used to alleviate the problem, where some model smoothness criteria are added to the data mismatch measure to form the overall optimization criterion [8], [9].

Even with the introduction of smoothness regularizers, the performance of deformable matching is sometimes still far from satisfactory, especially when the shape of the target object is highly deviated from the reference model and the level of outlier contamination is high. One possible direction to reduce the outlier influence is to enhance the model adequacy. For example, domain-specific constraints obtained via careful design can be imposed on shape variations for the particular application [13], [6]. Also, some model statistics obtained via learning can be incorporated to

- K.-W. Cheung is with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong.
E-mail: william@comp.hkbu.edu.hk.
- D.-Y. Yeung and R.T. Chin are with the Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. E-mail: {dyyeung, roland}@cs.ust.hk.

Manuscript received 1 Mar. 2000; revised 31 May 2001; accepted 14 Nov. 2001.

Recommended for acceptance by A. Kundu.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111593.