Disparity-Space Images and Large Occlusion Stereo

Stephen S. Intille and Aaron F. Bobick

Perceptual Computing Group
The Media Lab, Massachusetts Institute of Technology
20 Ames St., Cambridge MA 02139

Abstract. A new method for solving the stereo matching problem in the presence of large occlusion is presented. A data structure — the *disparity space image* — is defined in which we explicitly model the effects of occlusion regions on the stereo solution. We develop a dynamic programming algorithm that finds matches and occlusions simultaneously. We show that while some cost must be assigned to unmatched pixels, our algorithm's occlusion-cost sensitivity and algorithmic complexity can be significantly reduced when highly-reliable matches, or *ground control points*, are incorporated into the matching process. The use of ground control points eliminates both the need for biasing the process towards a smooth solution and the task of selecting critical prior probabilities describing image formation.

1 Introduction

Occluded regions are spatially coherent groups of pixels that can be seen in one image of a stereo pair but not in the other. These regions mark depth discontinuities and can be used to improve segmentation, motion analysis, and object identification processes, all of which must preserve object boundaries. There is psychophysical evidence that the human visual system uses geometrical occlusion relationships during binocular stereopsis[11] to reason about the spatial relationships between objects in the world. In this paper we present a stereo algorithm that does so as well.

Most stereo researchers have generally either ignored occlusion analysis or treated it as a secondary process that is postponed until matching is completed and smoothing is underway[6]. Consequently, occlusion regions are often a major source of error[3]. Stereo images of everyday scenes, such as people walking around a space, can contain contain disparity shifts and occlusion regions over eighty pixels wide[9] – much larger than occlusion regions found in typical stereo test imagery.

Our approach is to explicitly model occlusion edges and occlusion regions and to use them to drive the matching process. We develop a data structure which we will call the *disparity-space image* (DSI), and we use this data structure to develop a stereo algorithm that finds matches and occlusions *simultaneously*. We show that while some cost must be assigned to unmatched pixels, an algorithm's occlusion-cost sensitivity and algorithmic complexity can be significantly reduced

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

when highly-reliable matches, or *ground control points* (GCPs), are incorporated into the matching process.

Some previous stereo work has used the occlusion constraint explicitly in the matching process[1, 7, 5]. Our approach differs in that we use no additional criteria such as smoothness and intra and inter-scanline consistency since we use GCPs to eliminate sensitivity to occlusion costs.

2 The DSI Representation

The DSI is an explicit representation of matching space; it is related to figures that have appeared in previous work[4, 7, 13, 12]. We generate the DSI representation for i^{th} scanline in the following way: Select the i^{th} scanline of the left and right images, s_i^L and s_i^R respectively, and slide them across one another one pixel at a time. At each step, the scanlines are subtracted and the result is entered as the next line in the DSI. The DSI representation stores the result of subtracting every pixel in s_i^L with every pixel s_i^R and maintains the spatial relationship between the matched points. As such, it may considered an *(x, disparity)* matching space, with x along the horizontal, and disparity along the vertical. Given two images I_L and I_R the value of the DSI is given:

$$\mathrm{DSI}_i^R(x,d) = \begin{cases} I_R(x,i) - I_L(x+d,i) \\ \qquad \text{when } 0 \leq (x+d) < N \end{cases} \tag{1}$$

where all other values are not defined and $0 \leq d < N$ and $0 \leq x < N$. The superscript of R on DSI^R indicates the right DSI. DSI_i^L is simply a negated, skewed version of the DSI_i^R.

The above definition generates a "full" DSI where there is no limit on disparity. By considering camera geometry and some maximum possible disparity shift, we can crop the representation. Further, to make the DSI more robust to effects of noise, we use correlation with a simplified version of adaptive windows[10] that preserves sharp boundaries at occlusion jumps in the DSI_i^L [9].

Figure 1-c shows the cropped, correlation DSI for a scanline through the middle of the test image pair shown in Figure 1-b. Near-zero values have been enhanced. Notice the characteristic streaking pattern that results from holding one scanline still and sliding the other scanline across. When a textured region on the left scanline slides across the corresponding region in the right scanline, a line of matches can be seen in the DSI_i^L. When two textureless matching regions slide across each other, a diamond-shaped region of near-zero matches can be observed. The more homogeneous the region is, the more distinct the resulting diamond shape will be. The correct path through DSI space can be easily seen as a dark line connecting block-like segments.

3 Occlusion Analysis and DSI Path Constraints

In a discrete formulation of the stereo matching problem, any region with non-constant disparity must have associated unmatched pixels. Any slope or disparity jump creates blocks of occluded pixels. Because of these occlusion regions, the matching zero path through the image cannot be continuous. The regions labeled

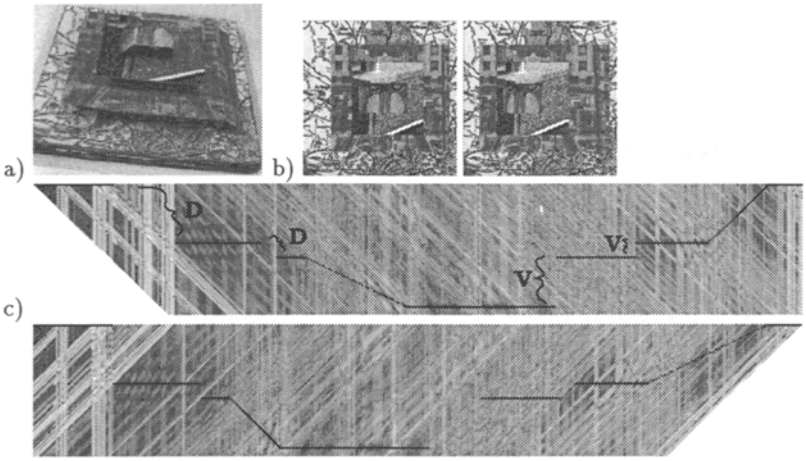

Fig. 1.: (a) A physical model of a sloping wedding cake, (b) a simulated image pair of the cake, (c) the enhanced, cropped DSI$_i^L$ and DSI$_i^R$ for one scan line.

"D" in Figure 1-c mark horizontal *gaps* in the enhanced zero line in DSI$_i^L$ and DSI$_i^R$. The regions labeled "V" mark vertical jumps from disparity to disparity. These jumps correspond to left and right occlusion regions. We use this "occlusion constraint" [7] to restrict the type of matching path that can be recovered from each DSI$_i^L$. Each time an occluded region is proposed, the recovered path is forced to have the appropriate vertical or diagonal jump.

Nearly all stereo scenes obey the *ordering constraint* (or *monotonicity constraint* [7]): if object a is to the left of object b in the left image then a will be to the left of b in the right image. Thin objects with large matching disparities violate this rule, but they are rare. By assuming the ordering rule we can impose a second constraint on the disparity path through the DSI that significantly reduces the complexity of the path-finding problem. In the DSI$_i^L$, moving from left to right, diagonal jumps can only jump forward (down and across) and vertical jumps can only jump backwards (up). In the DSI$_i^R$ the relationship is reversed: moving left to right diagonal jumps can only jump backwards and across and vertical jumps can only jump forwards (down). If this rule is broken the ordering constraint does not hold.

4 Finding the Best Path

Using the occlusion constraint and ordering constraint, the correct disparity path is highly constrained. From any location in the DSI$_i^L$, there are only three directions a path can take – a horizontal match, a diagonal occlusion, and a vertical occlusion. This observation allows us to develop a stereo algorithm that integrates matching and occlusion analysis into a single process.

Our algorithm for finding the best path through the DSI is formulated as a dynamic programming (DP) path-finding problem in $(x, \textit{disparity})$ space. We wish to find the minimum cost traversal through the DSI_i^L image when the occlusion constraints are imposed.

4.1 Dynamic Programming Constraints

The occlusion constraint and ordering constraint severely limit the direction the path can take from the path's current endpoint. If we base the decision of which path to choose at any pixel only upon the cost of each possible path we can take and not on any previous moves we have made, we satisfy the DP requirements and can use DP to find the optimal path.

Our DSI analysis led us to consider the occlusion problem in a "state-like" manner. As we traverse through the DSI image finding the optimal path, we can be in any of three states: *match (M), vertical occlusion (V), or diagonal occlusion (D).* Figure 2 symbolically shows the legal transitions between each type of state. The path is further constrained at the edges of the DSI image, where several types of transitions may be invalid.

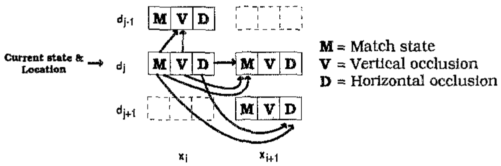

Fig. 2.: State diagram of moves the DP algorithm can choose through the DSI.

A cost is assigned to each pixel in the path depending upon the current state. We design our DP algorithm to *minimize* the cost of a path where the cost of a match is the value of the DSI_i^L pixel at the match point. The better the match, the lower the cost assessed. The algorithm will attempt to maximize the number of "good" matches in the final path. Since the algorithm will also propose unmatched points — occlusion regions — we need to assign a cost for unmatched pixels in the vertical or diagonal jumps. Otherwise the "best path" would be one that matches almost no pixels.

For the work presented here we chose a constant occlusion pixel cost. Without an additional constraint the algorithm is quite sensitive to this cost. In the next section we propose an alternative approach to reducing occlusion cost sensitivity that reduces complexity and does not artificially restrict the disparity path.

4.2 Ground Control Points

Unfortunately, slight variations in the occlusion pixel cost can change the globally minimum path through the DSI_i^L space, particularly with noisy data[5]. Because this cost is incurred for each proposed occluded pixel, the cost of a proposed occlusion region is linearly proportional to the width of the region.

In order to overcome this occlusion cost sensitivity, we need to impose another constraint in addition to the occlusion and ordering constraints. However, unlike previous approaches we do not want to bias the solution towards any generic property such as smoothness[7], inter-scanline consistency[12, 5], or intra-scanline "goodness"[5].

Instead, we use high confidence matching guesses: *Ground control points* (GCPs). These points are used to force the disparity path to make large disparity jumps that might otherwise have been avoided because of large occlusion costs.

Figure 3 illustrates this idea showing two GCPs and a number of possible paths between them. We note that regardless of which disparity path is chosen, the discrete lattice ensures that path-a, path-b, and path-c all require 6 occlusion pixels. Therefore, all three paths incur the same occlusion cost. Our algorithm will select the path that minimizes the cost of the proposed matches *independent of where occlusion breaks are proposed and the occlusion cost value.* If there is a single occlusion region between the GCPs in the original image, the path with the best matches is similar to path-a or path-b. On the other hand, if the region between the two GCPs is sloping gently, then a path like path-c, with tiny, interspersed occlusion jumps will be preferred. The path through *(x, disparity)* space, therefore, will be constrained solely by the occlusion and ordering constraints and the goodness of the matches between the GCPs. An exception to this situation occurs if the algorithm proposes additional occlusion regions as in path-d; such solutions typically have a much higher cost than the correct one.

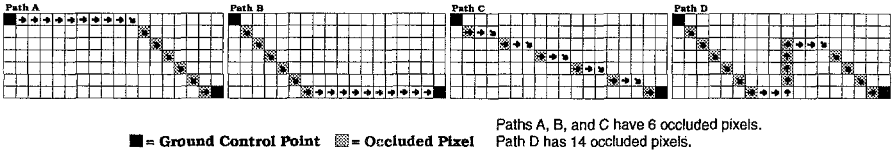
Path A
Path B
Path C
Path D
■ = Ground Control Point
▨ = Occluded Pixel
Paths A, B, and C have 6 occluded pixels.
Path D has 14 occluded pixels.

Fig. 3.: Four possible paths through two GCPs.

4.3 Selecting and Enforcing GCPs

If we force the disparity path through GCPs, their selection must be highly reliable. We use several heuristic filters to identify GCPs before we begin the DP processing. The first heuristic requires that a control point be both the best left-to-right and best right-to-left match[8]. Second, to avoid spurious "good" matches in occlusion regions, we also require that control points have match value that is smaller than the occlusion cost. Finally, to further reduce the likelihood of a spurious match, we exclude any proposed GCPs that have no immediate neighbors that are also marked as GCPs.

Once we have a set of control points, we force our DP algorithm to choose a path through the points by assigning zero cost for matching with a control point and a very large cost to every other path through the control point's column. In the DSI_i^L, the path must pass through each column at some pixel in some state. By assigning a large cost to all paths and states in a column other than a match at the control point, we have guaranteed that the path will pass through the point.

An important feature of this approach of incorporating GCPs is that this method allows us to have more than one GCP per column. Instead of forcing

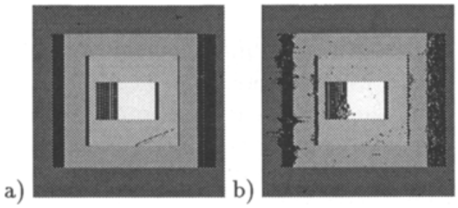

Fig. 4.: Results for the (a) noise-free and (b) noisy sloping wedding cake.

the path through one GCP, we force the path through one of a few GCPs. Even using multiple windows and left-to-right, right-to-left matching, it is still possible that we will label a GCP in error if only one per column is permitted. It is unlikely, however, that none of several proposed GCPs in a column will be the correct GCP. By allowing multiple GCPs per column, we have eliminated the risk of forcing the path through a point erroneously marked as high-confidence due image noise without increasing complexity or weakening the GCP constraint.

The use of GCPs reduces the complexity of the DP algorithm. With several GCPs the complexity can be less than 25% of the original problem[9].

5 Dynamic Programming Algorithm – Results

Input to our algorithm consists of a stereo pair. Epipolar lines are assumed to be known and corrected to correspond to horizontal scanlines. We assume that additive and multiplicative photometric bias between the left and right images is minimized, although the birch tree example shows our algorithm will work with significant additive differences.

The results generated by our algorithm for the noise-free sloping wedding cake are shown in Figure 4-a in the cyclopean view. The top layer of the cake has been shifted 84 pixels. Our algorithm found the occlusion breaks at the edge of each layer, indicated by black regions. Sloping regions have been recovered as a sloping region interspersed with tiny occlusion jumps. Since we have not used any smoothing or inter- or intra-scanline consistency, the solution in the sloping regions is governed only by the ground control points and the best matches in the region. Figure 4-b shows the results for the sloping wedding cake with noise (SNR = 18 dB). The algorithm still locates occlusion regions well.

Figure 5-a shows the "birch" image from the JISCT stereo test set[2]. The occlusion regions in this image are difficult to recover properly because of the skinny trees, some textureless regions, and a 15 percent brightness difference between images. The skinny trees make occlusion recovery particularly sensitive to occlusion cost when GCPs are not used, since there are relatively few good matches on each skinny tree compared with the size of the occlusion jumps to and from each tree. Figure 5-b shows the results of our algorithm *without* using GCPs. The occlusion cost prevented the path on most scanlines from jumping out to some of the trees. Figure 5-c shows the algorithm run with the same occlusion cost using GCPs. The occlusion regions around the trees are recovered reasonably well since GCPs on the tree surfaces eliminated the dependence on

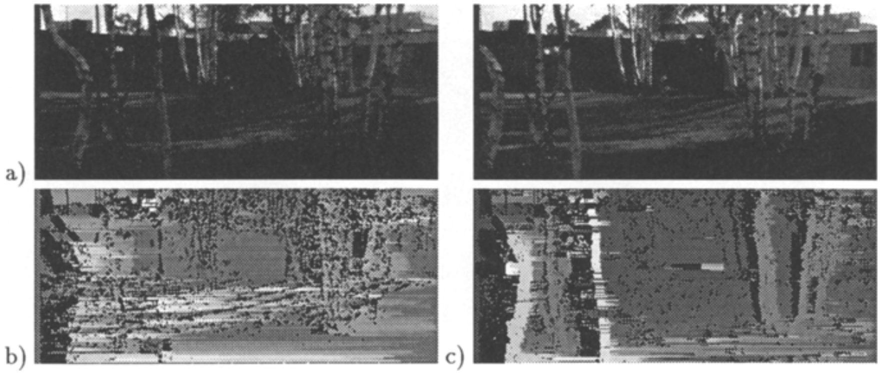

Fig. 5.: (a) Image pair and results without (b) and with (c) GCPs.

the occlusion cost. The algorithm fails where the image is washed-out, the image is textureless, or where no GCPs were recovered on some trees.

Figure 6-a is the left image of a stereo image pair of some people. Figure 6-b shows the left image results obtained by the algorithm developed by Cox *et al.*[5]. The Cox algorithm is a similar DP procedure which uses inter-scanline consistency instead of GCPs to reduce sensitivity to occlusion cost. Figure 6-c shows our results on the same image. The Cox algorithm does a reasonably good job at finding the major occlusion regions, although many rather large, spurious occlusion regions are proposed. When the algorithm generates errors, the errors are more likely to propagate over adjacent lines, since inter-and intra-scanline consistency are used[5]. To be able to find the numerous occlusions, the Cox algorithm requires a relatively low occlusion cost, resulting in false occlusions. Our higher occlusion cost and use of GCPs finds the major occlusion regions cleanly. For example, the man's head is clearly recovered by our approach. The algorithm did not recover the occlusion created by the man's leg as well as hoped since it found no good control points on the bland wall between the legs. The wall behind the man was picked up well by our algorithm, and the structure of the people in the scene is quite good. Most importantly, *we did not use any smoothness or inter- and intra-scanline consistencies to generate these results.*

We should note that our algorithm does not perform well on images that only have short match regions interspersed with many disparity jumps. In such imagery our conservative method for selecting GCPs fails to provide enough constraint to recover the proper surface. However, the results on the birch imagery illustrate that in real imagery with many occlusion jumps, there are likely to be enough stable regions to drive the computation.

6 Summary

We have presented a stereo algorithm that incorporates the detection of occlusion regions directly into the matching process. We develop an dynamic programming solution that obeys the occlusion and ordering constraints to find a best path through the disparity space image. To eliminate sensitivity to occlusion cost

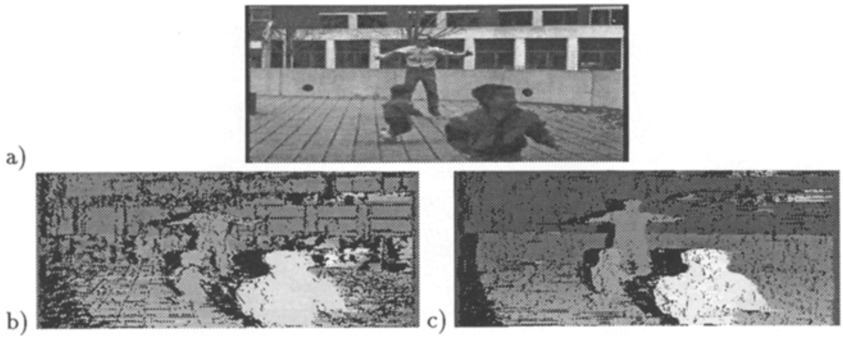

Fig. 6.: (a) Left image. Results of (b) Cox *et al.* algorithm[5], and (c) our algorithm.

we use ground control points (GCPs)— high confidence matches. These points improve results, reduce complexity, and minimize dependence on occlusion cost without arbitrarily restricting the recovered solution.

References

1. P. Belhumeur. Bayesian models for reconstructing the scene geometry in a pair of stereo images. In *Proc. Info. Sciences Conf.*, Johns Hopkins University, 1993.
2. R. Bolles, H. Baker, and M. Hannah. The JISCT stereo evaluation. In *Proc. Image Understanding Workshop*, pages 263–274, 1993.
3. R.C. Bolles and J. Woodfill. Spatiotemporal consistency checking of passive range data. SRI Technical Report – to be published, SRI International, September 1993.
4. S.D. Cochran and G. Medioni. 3-d surface description from binocular stereo. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 14(10):981–994, 1992.
5. I.J. Cox, S. Hingorani, B. Maggs, and S. Rao. Stereo without regularization. NEC Research Institute Report, NEC Research Institute, October 1992.
6. U.R. Dhond and J.K. Aggarwal. Structure from stereo – a review. *IEEE Trans. Sys., Man and Cyber.*, 19(6):1489–1510, 1989.
7. D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. In *Proc. European Conf. Comp. Vis.*, pages 425–433, 1992.
8. M.J. Hannah. A system for digital stereo image matching. *Photogrammetric Eng. and Remote Sensing*, 55(12):1765–1770, 1989.
9. S.S. Intille and A.F. Bobick. Disparity-space images and large occlusion stereo. MIT Media Lab Perceptual Computing Group Technical Report No. 220, Massachusetts Institute of Technology, October 1993.
10. T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. In *Proc. Image Understanding Workshop*, pages 383–389, 1990.
11. K. Nakayama and S. Shimojo. Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, 30(11):1811–1825, 1990.
12. Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 7:139–154, 1985.
13. Y. Yang, A. Yuille, and J. Lu. Local, global, and multilevel stereo matching. In *Proc. Comp. Vis. and Pattern Rec.*, 1993.