

Global Priors for Binocular Stereopsis

Peter N. Belhumeur

Center for Systems Science
Department of Electrical Engineering
Yale University
New Haven, CT 06520

Abstract

In this paper, we develop a Bayesian feedback method for incorporating global structure into prior models for binocular stereopsis. Since most stereo scenes contain either background continuation (large background surfaces continuing behind smaller foreground surfaces) or transparency continuation (small opaque patches on a transparent surface), highly non-local interactions are often present in the scene geometry. The commonly used local prior models which impose piecewise smoothness constraints on the reconstructions do not capture the probabilistic subtleties of global 3-D structures. Therefore, we develop a hybridized prior which balances the local properties of the scene geometry with the global properties. Experimental results demonstrating the potential of this technique are provided.

1 Introduction

In Bayesian approaches to binocular stereopsis, most prior models share the same drawback: they have no way of modeling the global structures in the scene geometry. It is common to assume the disparity estimates are generated by Markov processes – processes with the property that the conditional distribution of the future, given the present states of the process, does not depend on the past (see [2]). While one can develop increasingly complicated prior models using the techniques of Gibbs' distributions and Markov random fields (see for instance [6, 8, 5, 11, 15, 3, 4, 7, 12, 9]), these models are limited in that they only allow local interactions of the random variables.

As an example consider the following Gibbs' distribution for modeling the joint probability of the disparity values along an epipolar line.

$$P(\mathbf{d} \mid F_l, F_r) = \frac{1}{Z} e^{-E_D - E_P} \quad (1)$$

where

$$E_D = \sum_{i \in O} (F_l(x_i + d_i) - F_r(x_i - d_i))^2 + \sum_O \alpha_O$$

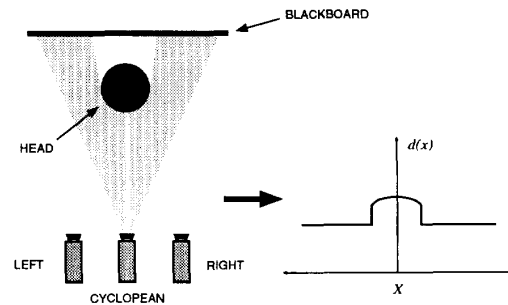


Figure 1: **Background continuation:** The figure illustrates a horizontal cross section of a spherical head in front of flat blackboard, as well as a graph of the corresponding disparity function $d(x)$. A Markov process can not model the probabilistic subtleties of continuing background surface.

$$E_P = \lambda^2 \sum_{i \in B \cup \{n\}} (d_{i+1} - d_i)^2 + \sum_B \alpha_B$$

and where $\{x_0, \dots, x_i, \dots, x_n\}$ are n evenly spaced sample points along a cyclopean epipolar line X ; $\mathbf{d} = \{d_0, \dots, d_i, \dots, d_n\}$ are disparity values corresponding to the points $\{x_0, \dots, x_i, \dots, x_n\}$; I is the index set $\{1, \dots, n\}$; $O \subset I$ marks the points x_i which are occluded; $B \subset I$ marks the points x_i such that x_i and x_{i+1} are on different objects; F_l and F_r are functions representing the features along the left and right epipolar lines; and α_O , λ , and α_B are preset constants. (This model is developed in detail in [1].)

Note that, according to the model in Eq. 1, the most likely disparity is the set $\mathbf{d}$ which maximizes $P(\mathbf{d} \mid F_l, F_r)$ or, equivalently, minimizes $E_D + E_P$. The prior term E_P successfully models the fact that along the surface of an object there is often little variation in disparity, but across the boundary of an object the disparity jumps discontinuously. Yet, this prior model had several shortcomings. Most notably, since occluded objects tend to disappear and reappear behind foreground objects, a phenomenon that we call "background continuation," global interactions must

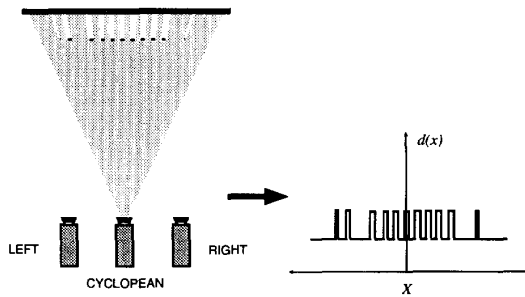


Figure 2: **Transparent continuation – overhead view:** The figure shows, on the left, a horizontal cross section of flat wall viewed through a store-front window with lettering, and, on the right, a graph of the corresponding disparity function $d(x)$. A Markov process can not model the probabilistic subtleties of transparent surfaces.

also be present.¹ Look around you; in almost any scene you will see background objects continuing behind smaller foreground objects.

As an example, Fig. 1 shows a horizontal cross section of a spherical head in front of flat blackboard, viewed by a left, right, and cyclopean camera. For this scene the corresponding disparity function $d(x)$ is divided by the boundaries of the foreground sphere into three piecewise smooth regions. Although Markov prior models implicitly assume that the disparity in these regions is independent, we see from the figure that this is clearly incorrect: the fact that the disparity is the same for the planar regions to the left and right of the foreground sphere is not a coincidence, but rather a result of the global structure of the scene.

A closely related phenomenon to background continuation is what we call “transparent continuation.” Here we have foreground objects which in some regions are opaque and in other regions transparent. (These types of scenes have proven to be difficult for computational stereo algorithms, even though the human visual system handles these correctly.)

As an example, imagine looking through the lettering of a storefront window at a wall in the background. The letters of the words are opaque, but the glass sign on which they are written is transparent. If we take a horizontal cross section of this scene (see Fig. 2), we see the corresponding disparity function $d(x)$ is divided by the boundaries of the foreground letters into many piecewise linear regions – some regions corresponding to the foreground letters and the others to the background wall. As before the Markov prior model in Eq. 1 implicitly assumes that the disparity in these regions is independent, yet we see from the figure that this is again incorrect: the fact that the disparity clusters into two distinct lines is not a coincidence, but rather a result of the global structure

¹We use the term “global” to mean highly non-local, i.e. existing on the scale greater than one-tenth of observed scene.

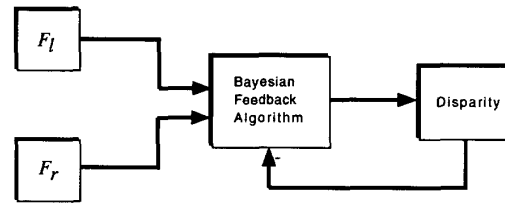


Figure 3: **Feedback Schematic:** In the feedback scheme, the previous estimate is analyzed for global structure and the results are used to influence the current estimate by adjusting specific terms in the prior model.

of the scene. (For more discussion of transparency see [10].)

It seems that local prior models can not properly model stereo scenes of this type. Thus, previous Bayesian approaches to binocular stereopsis [11, 15, 3, 7, 12] made no consideration of global structure. (Although [14] does not have explicit mechanisms for handling transparency and background continuation, the disparity gradient limit does seem to implicitly accommodate for these types of scenes.) What is needed is a prior model that balances local smoothness with global structure.

In this paper we argue that to incorporate global interactions into the prior model, we can develop a Bayesian feedback algorithm in which the prior is iteratively adapted using the previous estimate. Here the previous estimate is analyzed for global structure, and the results are used to influence the current estimate by adjusting specific terms in the prior. Figure 3 shows a simple schematic diagramming the flow of the feedback algorithm.

2 Deriving the Feedback Prior

The Markov prior for the Bayesian model in Eq. 1 can be written as

$$P(\mathbf{d}) = \frac{1}{Z_P} e^{-E_P}.$$

In this section, we develop an alternative feedback prior $P(\mathbf{d}^k | \mathbf{d}^{k-1})$ which biases in favor of global structure. Yet, what considerations should be made in choosing the new prior model? How are we to incorporate long range interactions of the random variables without ignoring the strengths of Markov priors? To do this, we introduce a hybridized model which balances the local prior in Eq. 1 with the global priors developed here.

Assuming independence of the local and global priors, we write our feedback prior as

$$P(\mathbf{d}^k | \mathbf{d}^{k-1}) = P_L(\mathbf{d}^k) P_G(\mathbf{d}^k | \mathbf{d}^{k-1})$$

where $P_L(\mathbf{d}_k)$ is local smoothness priors and $P_G^k(\mathbf{d}_k | \mathbf{d}_{k-1})$ is a dynamic, global prior which is

adapted after each iteration. For the purpose of our discussion, let us simplify the development and look only at the disparity along a single epipolar line. (Naturally, all of these ideas generalize to 2-D.) For evenly spaced samples of the disparity along a single epipolar line, we can write the global prior

$$P_G(\mathbf{d}^k | \mathbf{d}^{k-1}) = P_G(\mathbf{d}^k | \mathbf{d}^{k-1})$$

where $\mathbf{d}^k$ is the vector of sampled disparities at iteration k . Finally, let us assume that, given the vector of the previous disparity estimates $\mathbf{d}^{k-1}$, the current disparity estimates for the global prior are all independent. So we rewrite

$$P_G(\mathbf{d}^k | \mathbf{d}^{k-1}) = \prod_{i=1}^n P_G(d_i^k | \mathbf{d}^{k-1})$$

where d_i^k is the disparity estimate at point x_i for the current iteration.

Empirically, we have found that for the global prior to be effective, it needs to satisfy three criteria. First, the global structures it looks for must have a simple representation and be common to most images. Second, the influence of the random variables of the global prior can not be truly global (i.e. the interaction of the random variables is independent of the spatial distance separating them), but rather must be localized. By this we mean that the influence of the random variables in the global prior should fall off with the spatial distance separating them. Finally, the prior should give a level of confidence or probability (as opposed to a yes or no), as to whether a point lies on a particular global structure.

An Example: Planar Structure

Before we can choose $P_G(d_i^k | \mathbf{d}^{k-1})$, we must know what sort of global structures to look for in the scene geometry. Note that for both of the above mentioned scenes, a good portion of the surfaces are clustered into common depth planes. In the scene from Fig. 1, all of the background points lie in the plane of the blackboard. And in the scene from Fig. 2, this phenomenon is even more exaggerated: every point in the scene lies in one of two distinct planes. Note also that for these scenes, the clustering exists even though many of the points are separated by either foreground occluding objects or transparent regions. In fact, due to background and transparency continuation, this type of planar clustering is present (at some scale) in most stereo images (see also [13]). (Because we restrict our discussion to disparity estimates along a single epipolar line, this planar clustering reduces to linear clustering.)

Therefore, we develop a global prior to accomplish precisely what our local prior could not – to look for and bias in favor of linear structures which may be interrupted by either foreground occlusion or transparent regions.² To do this, we develop a measure

²Naturally, this method could be generalized to include other structures, both parametric and non-parametric.

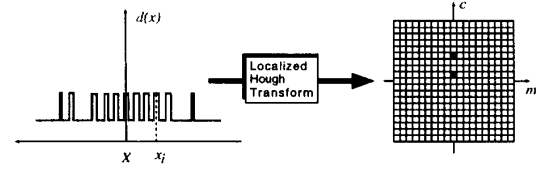


Figure 4: **Localized Hough space array:** The figure shows the localized Hough space histogram (LHSH) for a point along the epipolar line in Fig. 2. On the left of the figure is the disparity function, with the center point for the LHSH marked. On the right of the figure is the corresponding LHSH. The gray bin corresponds to the foreground letters; the black bin corresponds to the patterned wall.

$P_G(d_i^k | \mathbf{d}^{k-1})$ that biases the disparity d_i^k to lie on one of the linear structures found in the previous disparity estimates $\mathbf{d}^k$. While there are many methods that one might use to construct $P_G(d_i^k | \mathbf{d}^{k-1})$, we use a variation on the standard Hough transform method for detecting lines.³

In the standard Hough method, one transforms 2-D search space into a space tuned to detecting straight lines. To do this, note that a point (x_j, d_j) in the disparity space has all of the lines $mx_j + c = d_j$ passing through it. These lines correspond to the single line $x_j m - d_j = -c$ in the Hough space. So for each sampled disparity d_j along the epipolar line, we draw a line in Hough space. If divide up our Hough space into square bins, then we can count the number of lines which are drawn through each bin. (We call this 2-D array of bins the Hough space histogram.) After we have done this for every sampled disparity, the size of the count in each bin reflects the strength of presence of the corresponding line in disparity space. While this technique is effective at finding linear structures in the disparity space, it is not spatially localized (i.e. all points in disparity space contribute equally to the Hough space histogram).

We improve on this technique by constructing what we call a “localized Hough transform” and the corresponding localized Hough space histogram $H_i[m][c]$ (LHSH) at each sampled point x_i along the epipolar line. To construct $H_i[m][c]$ at a point x_i , we draw, as before, a line in Hough space for each sampled disparity d_j (at the point x_j) along the epipolar line. However, we weight the counts in each bin by an amount w_j which falls off with the distance $|x_j - x_i|$. Thus, the farther a sampled point x_j is from the transform’s center point x_i , the less weight it is given in the localized Hough space. We take the weights to be given by the normal distribution

$$w_j = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_j - x_i)^2 / 2\sigma^2}$$

³While Hough transform methods have come under fire of late, few researchers are aware of their relation to maximum likelihood estimation (MAP).

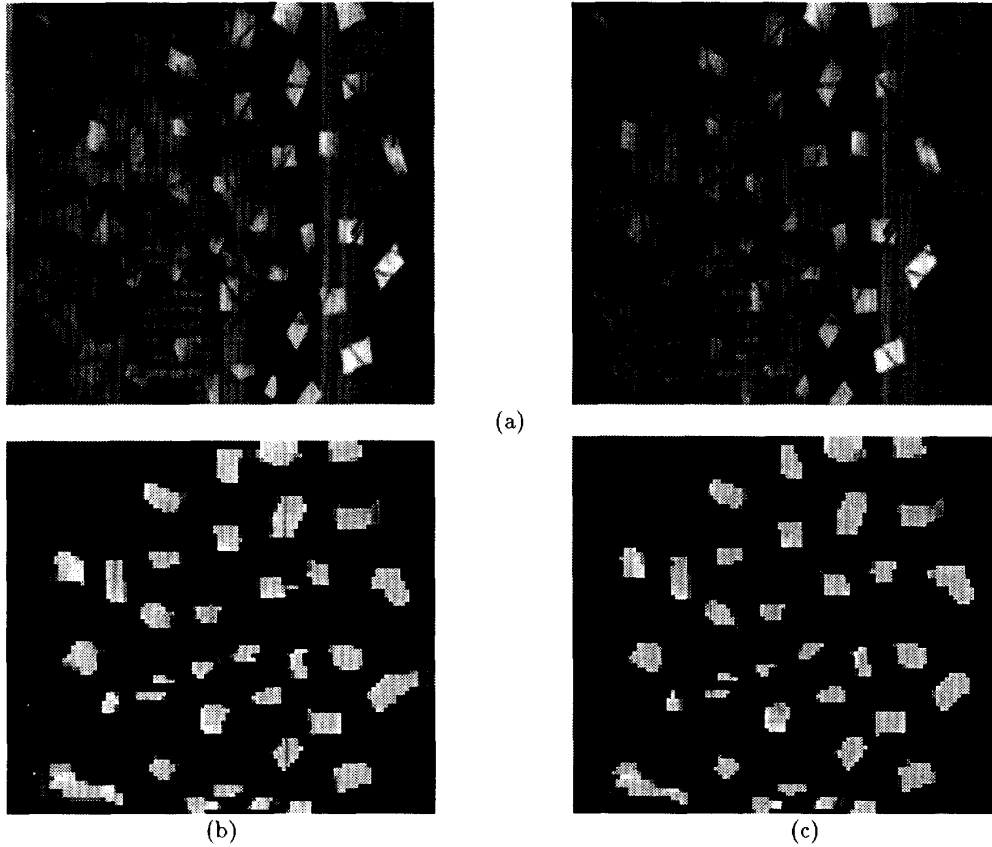


Figure 5: **Result:** (a) Transparent surface – left and right images. (b) Image of depth without using feedback. (c) Image of depth using feedback.

where σ is inversely proportional to the rate of falloff.

As an example, Fig. 4 shows, on the left, the disparity space for an epipolar line in the transparency scene from Fig. 2, and, on the right, the LSHH corresponding to the marked point. The black bin corresponds to the place of intersection of all the lines drawn for points sampled from the background wall. The gray bin corresponds to the place of intersection of all of the lines drawn for points sampled from the foreground letters. (So in this figure, dark corresponds to more counts in the Hough space array, and light corresponds to less.) As might be expected, the LSHH identifies the two linear disparity clusters, even though they are interrupted by transparent regions.

From the LSHH, we wish to construct our density for the global prior. We take the simplest route and define the density to be the normalized histogram

$$P_G(d_i^k | \mathbf{d}^{k-1}) = \frac{\sum_{\Lambda} H_i^k[m][c]}{\sum_{m,c} H_i^k[m][c]}$$

where $\Lambda = \{(m, c) | (x_i, d_i) \text{ is on the line } (m, c)\}$, and $H_i^k[m][c]$ is the LSHH for the k th iteration.

3 The Feedback Model with Results

In the last section, we derived the feedback prior model. In this section, we write down the resulting posterior distribution and present results demonstrating its effectiveness. The new posterior distribution is given as

$$P(\mathbf{d} | F_l, F_r) = \frac{1}{Z} e^{-E_D - E_P} \quad (2)$$

where E_D is as previously stated, but where E_P is the sum of the local and global prior terms. Specifically,

$$E_P = E_{P_L} + E_{P_G}$$

where E_{P_L} is the prior from Eq. 1 and, ignoring constants,

$$E_{P_G} = -\log \sum_{\Lambda} H_i^k[m][c].$$

To illustrate the effectiveness of this feedback method, we choose a stereo pair with transparency. Figure 5a shows a stereo pair of small pieces of paper

stuck to a transparent Plexiglas window. The window is placed in front of a planar background which falls away from the viewer as you move from left to right.

When the stereogram is fused, the viewer clearly sees the pieces of paper floating *within the same invisible plane*, several inches in front of the background plane. Figure 5b contains the results obtained by maximizing the posterior distribution in Eq. 1. In this result no feedback was used. While the result is fairly accurate, notice that in the transparent regions between many of the foreground pieces of paper, the depth reconstruction degenerates. This negative result is due to the fact that there is not enough information in the small background regions to independently determine the depth. And, the local priors are unable to propagate the background depths into the small transparent regions between the pieces of paper. However, Fig. 5c contains the results obtained by maximizing feedback posterior distribution in Eq. 2. Here we use the same local prior as before. Note that, except in two or three places, this feedback corrects the previous mistakes.

References

- [1] P. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *Proc. 4th Int. Conf. on Computer Vision*, Berlin, Germany, 1993.
- [2] R. Bhattacharya and E. Waymire. *Stochastic Processes with Applications*. John Wiley & Sons, New York, 1990.
- [3] B. Cernuschi-Frias, D. Cooper, Y. Hung, and P. Belhumeur. Toward a model-based bayesian theory for estimating and recognizing parameterized 3-d objects using two or more images taken from different positions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 1028–1052, November 1989.
- [4] J. Clark and A. Yuille. *Data Fusion for Sensory Information*. Kluwer Academic Press, Boston, 1990.
- [5] F. Cohen, D. Cooper, J. Silverman, and E. Hinkle. Simple parallel hierarchical relaxation algorithms for segmenting textured images based on noncausal markovian random field models. In *Proc. 7th Int. Conf. on Pattern Recognition*, pages 1104–1107, Montreal, Canada, 1984.
- [6] G. Cross and A. Jain. Markov random field texture models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.
- [7] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: surface reconstruction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(5):401–412, 1991.
- [8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [9] Z. Kato, M. Berthold, and J. Zerubia. Multiscale markov random field models for parallel image classification. In *Proc. 4th Int. Conf. on Computer Vision*, pages 253–257, 1993.
- [10] S. Madarasmí, D. Kersten, and T. Pong. *The computation of stereo disparity for transparent and for opaque surfaces*. Morgan Kaufmann Publishers, Giles, C.L., Hanson, S.J., and Cowan, J.D. (Eds.), 1993.
- [11] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solutions of ill-posed problems in computational vision. *J. of the Am. Stat. Soc.*, 82(397):76–89, 1987.
- [12] L. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *Int. Journal of Computer Vision*, 8(1):71–91, July 1992.
- [13] M. Nitzberg and D. Mumford. The 2.1d sketch. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 138–144, 1990.
- [14] S. Pollard, J. Mayhew, and J. Frisby. PMF:A Stereo Correspondence Algorithm Using A Disparity Gradient. *Perception*, 14:449–470, 1985.
- [15] R. Szeliski. *A Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Press, Boston, 1989.