The Computation of Stereo Disparity for Transparent and for Opaque Surfaces

Suthep Madarasmi
Computer Science Department
University of Minnesota
Minneapolis, MN 55455

Daniel Kersten
Department of Psychology
University of Minnesota

Ting-Chuen Pong
Computer Science Department
University of Minnesota

Abstract

The classical computational model for stereo vision incorporates a uniqueness inhibition constraint to enforce a one-to-one feature match, thereby sacrificing the ability to handle transparency. Critics of the model disregard the uniqueness constraint and argue that the smoothness constraint can provide the excitation support required for transparency computation. However, this modification fails in neighborhoods with sparse features. We propose a Bayesian approach to stereo vision with priors favoring cohesive over transparent surfaces. The disparity and its segmentation into a multi-layer "depth planes" representation are simultaneously computed. The smoothness constraint propagates support within each layer, providing mutual excitation for non-neighboring transparent or partially occluded regions. Test results for various random-dot and other stereograms are presented.

1 INTRODUCTION

The horizontal disparity in the projection of a 3-D point in a parallel stereo imaging system can be used to compute depth through triangulation. As the number of points in the scene increases, the correspondence problem increases in complexity due to the matching ambiguity. Prior constraints on surfaces are needed to arrive at a correct solution. Marr and Poggio [1976] use the smoothness constraint to resolve matching ambiguity and the uniqueness constraint to enforce a 1-to-1 match. Their smoothness constraint tends to oversmooth at occluding boundaries and their uniqueness assumption discourages the computation of stereo transparency for two overlaid surfaces. Prazdny [1985] disregards the uniqueness inhibition term to enable transparency perception. However, their smoothness constraint is locally enforced and fails at providing excitation for spatially disjoint regions and for sparse transparency.

More recently, Bayesian approaches have been used to incorporate prior constraints (see [Clark and Yuille, 1990] for a review) for stereopsis while overcoming the problem of oversmoothing. Line processes are activated for disparity discontinuities to mark the smoothness boundaries while the disparity is simultaneously computed. A drawback of such methods is the lack of an explicit grouping of image sites into piece-wise smooth regions. In addition, when presented with a stereogram of overlaid (transparent) surfaces such as in the random-dot stereogram in figure 5, multiple edges in the image are obtained while we clearly perceive two distinct, overlaid surfaces. With edges as output, further grouping of overlapping surfaces is impossible using the edges as boundaries. This suggests that surface grouping should be performed simultaneously with disparity computation.

2 THE MULTI-LAYER REPRESENTATION

We propose a Bayesian approach to computing disparity and its segmentation that uses a different output representation from the previous, edge-based methods. Our representation was inspired by the observations of Nakayama *et al.* [1989] that mid-level processing such as the grouping of objects behind occluders is performed for objects within the same "depth plane".

As an example consider the stereogram of a floating square shown in figure 1a. The edge-based segmentation method computes the disparity and marks the disparity edges as shown in figure 1b. Our approach produces two types of output at each pixel: a layer (depth plane) number and a disparity value for that layer. The goal of the system is to place points that could have arisen from a single smooth surface in the scene into one distinct layer. The output for our multi-surface representation is shown in figure 1c. Note that the floating square has a unique layer label, namely layer 4, and the background has another label of 2. Layers 1 and 3 have no data support and are, therefore, inactive.

The rest of the pixels in each layer that have no data support obtain values by a membrane fitting process using the computed disparity as anchors. The occluded parts of surfaces are, thus, represented in each layer. In addition, disjoint regions of a single surface due to occlusion are represented in a single layer. This representation of occluded parts is an important difference between our representation and a similar representation for segmentation by Darrell and Pentland [1991].

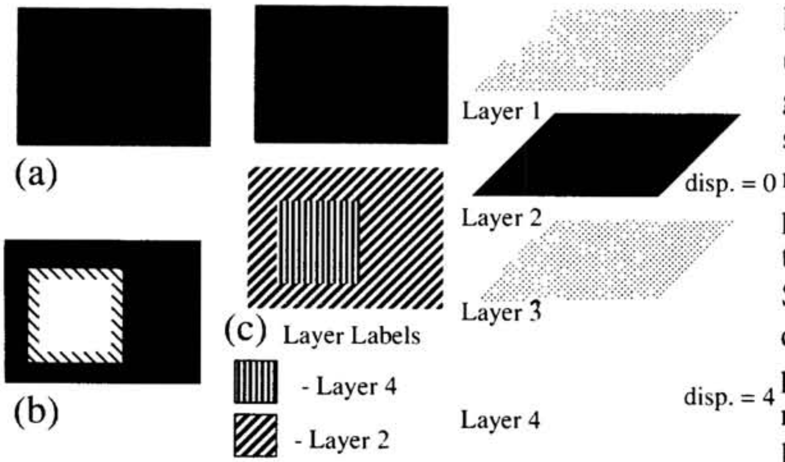

Figure1: a) A gray scale display of a noisy stereogram depicting a floating square. b. Edge based method: disparity computed and disparity discontinuity computed. c. Multi-Surface method: disparity computed, surface grouping performed by layer assignment, and disparity for each layer filled in.

3 ALGORITHM AND SIMULATION METHOD

We use Bayes' [1783] rule to compute the scene attribute, namely disparity u and its layer assignment l for each layer:

$$p(u, l|d^L, d^R) = \frac{p(d^L, d^R|u, l)p(u, l)}{p(d^L, d^R)}$$

where d^L and d^R are the left and right intensity image data. Each constraint is expressed as a local cost function using the Markov Random Field (MRF) assumption [Geman and Geman, 1984], that pixels values are conditional only on their nearest neighbors. Using the Gibbs-MRF equivalence, the energy function can be written as a probability function:

$$p(x) = \frac{1}{Z} e^{-\frac{E(x)}{T}}$$

where Z is the normalizing constant, T is the temperature, E is the energy cost function, and x is a random variable

Our energy constraints can be expressed as

$$E = \lambda_D V_D + \lambda_S V_S + \lambda_G V_G + \lambda_E V_E + \lambda_R V_R$$

where the λ's are the weighting factors and the V_D,V_S, V_G, V_E, V_R functions are the data matching cost, the smoothness term, the gap term, the edge shape term, and the disparity versus intensity edge coupling term, respectively.

The data matching constraint prefers matches with similar intensity and contrast:

$$V_D = \sum_{i}^{M} \left[|d_i^R - d_k^L| + \gamma \sum_{j \in N_i} |(d_j^R - d_i^R) - (d_m^L - d_k^L)| \right]$$

with the image indices k and m given by the ordered pairs $k = (row(i), col(i)+u_{C,i})$, $m = (row(j), col(j) + u_{C_i i})$, M is the number of pixels in the image, C_i is the layer classification for site i, and u_{l_i} is the disparity at layer l. The γ weighs absolute intensity versus contrast matching.

The λ_D is higher for points that belong to unambiguous features such as straight vertical contours, so that ambiguous pixels rely more on their prior constraints.

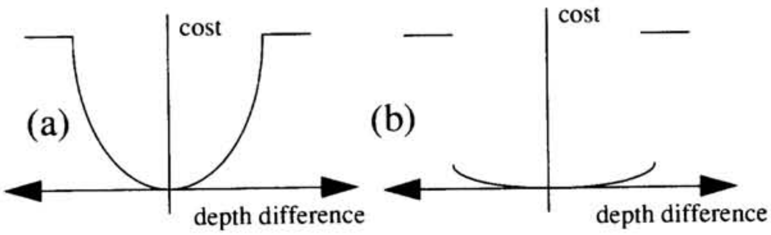

Figure 2: Cost function V_S. a) The smoothness cost is quadratic until the disparity difference is high and an edge process is activated. b) In our simulations we use a threshold below which the smoothness cost is scaled down and above which a different layer assignment is accepted at a constant high cost.

Also, if neighboring pixels have a higher disparity than the current pixel and are in a different layer, its λ_D is lowered since its corresponding point in the left image is likely to be occluded.

The equation for the smoothness term is given by:

$$V_S = \sum_{i}^{M} \sum_{l}^{L} \sum_{j \in N_i} V_s(u_{li}, u_{lj}) a_l$$

where, N_i are the neighbors of i, V_s is the local smoothness potential, a_l is the activity level for layer l defined by the percent of pixels belonging to layer l, and L is the number layers in the system. The local smoothness potential is given by:

$$V_s = \sigma_1(u_{li}, u_{lj}) + \mu \sum_{k} \sigma_2(\Delta_k u_{li}, \Delta_k u_{lj}); \quad \sigma_n(a,b) = \begin{cases} \frac{(a-b)^2}{\beta_n} & \text{if } (a-b)^2 < T_n \\ 1 & \text{otherwise} \end{cases}$$

where μ is the weighting term between depth smoothness and directional derivative smoothness. The Δ_k is the difference operation in various directions k, and T is the threshold. Instead of the commonly used quadratic smoothness function graphed in figure 2a, we use the σ function graphed in figure 2b which resembles the Ising potential. This allows for some flexibility since λ_S is set rather high in our simulations.

The V_G term ensures a gap in the values of corresponding pixels between layers:

$$V_G = \sum_{i}^{M} \sum_{l \neq C_i}^{L} \sum_{j \in N_i} V_g(u_{C_i i}, u_{lj}) a_l a_{C_i}; \quad V_g(u_{C_i i}, u_{lj}) = \begin{cases} 0 & \text{if } |u_{C_i i} - u_{lj}| \geq T \\ 1 & \text{otherwise} \end{cases}$$

This ensures that if a site i belongs to layer C_i, then all points j neighboring i for each layer l must have different disparity values u_{lj} than $u_{C_i i}$.

The edge or boundary shape constraint V_E incorporates two types of constraints: a cohesive measure and a saliency measure. The costs for various neighborhood configurations are given in figure 3.

The constraint V_R ensures that if there is no edge in intensity then there should be no edge in the disparity. This is particularly important to avoid local minima for gray scale images since there is so much ambiguity in the matching.

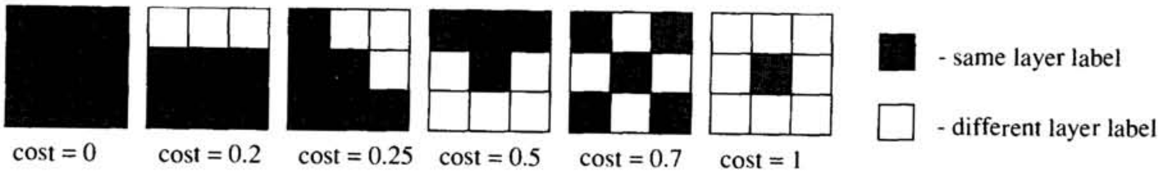

Figure 3: Cost function V_E. The costs associated nearest neighborhood layer label configurations. a) Fully cohesive region (lowest cost) b) Two opaque regions with straight line boundary. c) Two opaque regions with diagonal line boundary. d) Opaque regions with no figural continuity. e) Transparent region with dense samplings. f) Transparent region with no other neighbors (highest cost).

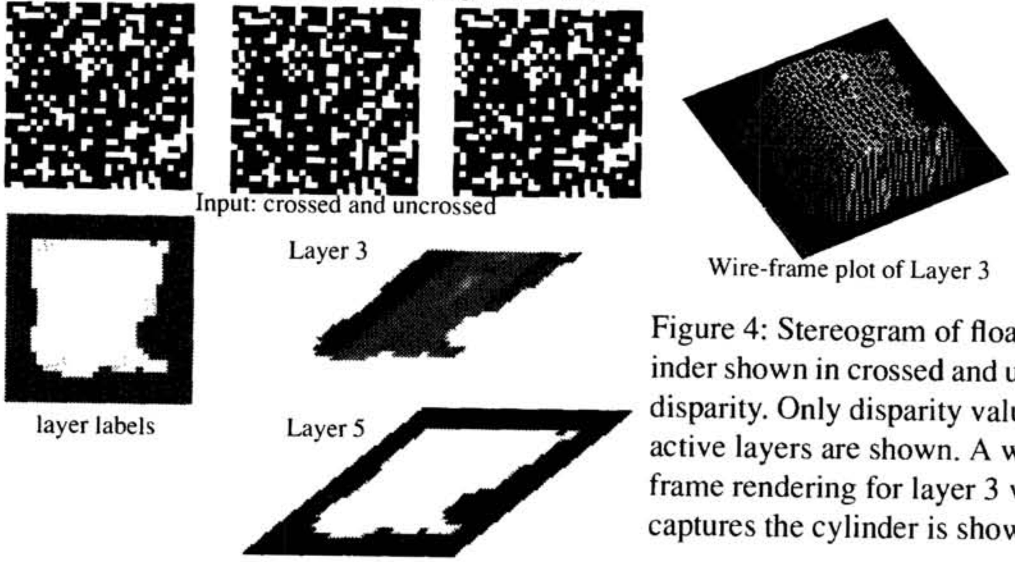

Figure 4: Stereogram of floating cylinder shown in crossed and uncrossed disparity. Only disparity values in the active layers are shown. A wire-frame rendering for layer 3 which captures the cylinder is shown.

The Gibbs Sampler [Geman and Geman, 1984] with simulated annealing is used to compute the disparity and layer assignments. After each iteration of the Gibbs Sampler, the missing values within each layer are filled-in using the disparity at the available sites. A quadratic energy functional enforces smoothness of disparity and of disparity difference in various directions. A gradient descent approach minimizes this energy and the missing values are filled-in.

4 SIMULATION RESULTS

After normalizing each of the local costs to lie between 0 and 1, the values for the weighting parameters used in decreasing order are: $\lambda_S, \lambda_R, \lambda_D, \lambda_E, \lambda_G$ with the λ_D value moved to follow λ_G if a pixel is partially occluded. The results for a random-dot stereogram with a floating half-cylinder are shown in figure 4. Note that for clarity only the visible pixels within each layer are displayed, though the remaining pixels are filled-in. A wire-frame rendering for layer 3 is also provided.

Figure 5 is a random-dot stereogram with features from two transparent fronto-parallel surfaces. The output consists primarily of two labels corresponding to the foreground and the background. Note that when the stereogram is fused, the percept is of two overlaid surfaces with various small, noisy regions of incorrect matches.

Figure 6 is a random-dot stereogram depicting many planar-parallel surfaces. Note

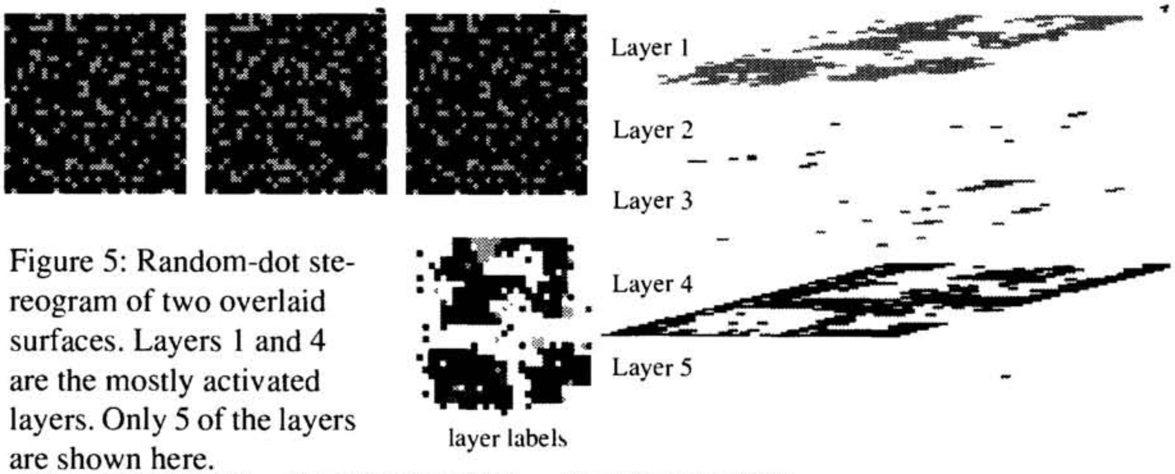

Figure 5: Random-dot stereogram of two overlaid surfaces. Layers 1 and 4 are the mostly activated layers. Only 5 of the layers are shown here.

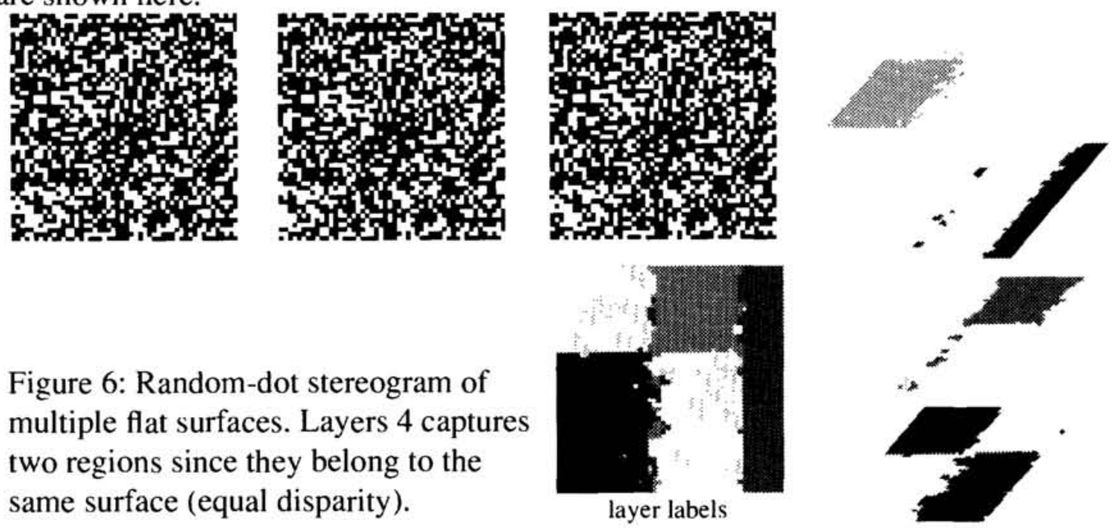

Figure 6: Random-dot stereogram of multiple flat surfaces. Layers 4 captures two regions since they belong to the same surface (equal disparity).

that there are two disjoint regions which are classified into the same layer since they form a single surface.

A gray-scale stereogram depicting a floating square occluding the letter 'C' also floating above the background is shown in figure 7. A feature-based matching scheme is bound to fail here since locally one cannot correctly attribute the computed disparity at a matched corner of the rectangle, for example, to either the rectangle, the background, or to both regions. Our V_R constraint forces the system to attempt various matches until points with no intensity discontinuity have no disparity discontinuity. Another important feature is that the two ends of the letter 'C' are in the same "depth plane" [Nakayama *et al.*, 1989] and may later be merged to complete the letter.

Figure 8 is a gray scale stereogram depicting 4 distant surfaces with planar disparity. At occluding boundaries, the region corresponding to the further surface in the right image has no corresponding region in the left image. A high λ_D would only force these points to find an incorrect match and add to the systems errors. The λ_D reduction factor for partially occluded points reduces the data matching requirement for such points. This is crucial for obtaining correct matches especially since the images are sparsely textured and the dependence on accurate information from the textured regions is high.

A transparency example of a fence in front a bill-board is given in figure 9. Note