

# **Inference of Segmented Overlapping Surfaces From Binocular and Multiple-View Stereo**

**Philippos Mordohai, Gerard Medioni, Mi-Suen Lee**

## **Affiliation of Authors**

- Philippos Mordohai is with the Institute for Robotics and Intelligent Systems, and with the Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-0273. E-mail: mordohai@iris.usc.edu.
- Gerard Medioni is with the Institute for Robotics and Intelligent Systems, and with the Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-0273. E-mail: medioni@iris.usc.edu.
- Mi-Suen Lee is with Philips Research, Philips Electronics North America Corporation, 345 Scarborough Road, Briarcliff Manor, NY 10510. E-mail: msl@philabs.research.philips.com.

Note: Part of this research that refers to binocular stereo was published in the Proceedings of CVPR 1998 as: M.S. Lee and G. Medioni, Inferring Segmented Surface Description from Stereo Data, Proc. CVPR, pp. 346-352, 1998.

## **Abstract**

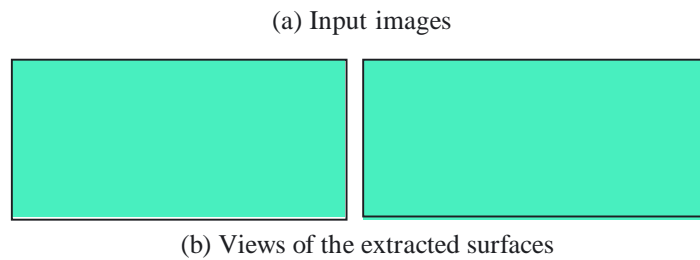
We present an integrated approach to the derivation of scene descriptions from two or more stereo images, where evaluation of feature correspondences and grouping into perceptual structures are addressed within the same framework. Special attention is given to the development of a methodology with general applicability. In order to handle the issues of noise, lack of image features, surface discontinuities and half occluded regions, we adopt a tensor representation for the data, and introduce a robust computational technique called tensor voting for information propagation. The major contributions of our approach are the fact that all processing is done in 3-D, the choice of saliency instead of cross-correlation as the criterion for determining the correctness of correspondences, and the integration of feature matching with surface and curve extraction. We have obtained promising results on synthetic and real data, as well as random dot stereograms.

## **Principles of our approach at Binocular Stereo**

We will first briefly describe our approach at the case of binocular stereo [1], and then show how it can be extended to multiple views. Given that the problem of binocular stereo is ill-posed, several constraints that should be imposed on the solution have been proposed. Besides the epipolar constraint, the most widely used among these constraints include the continuity constraint, the uniqueness constraint and the ordering constraint ([3], [4], [6]). Our framework allows us to impose the smoothness constraint locally in three dimensions, rather than one or two. It makes use of the epipolar constraint by forcing matches to only occur along these lines, but computes saliency across epipolar lines. It enforces uniqueness of matches locally, but allows this constraint to be overridden globally, leading to the inference of overlapping regions. We do

not impose the ordering constraint explicitly because it is violated by thin or small objects, transparent surfaces and acute concavities.

The steps from the image pair to the description of the scene in terms of perceptual structures, surfaces and curves, are the following. First points of interest, or features, are extracted from each image. These are pixels with non-zero intensity variance within a small window. The next step is feature matching between two images, along corresponding epipolar lines. We retain all local maxima of cross-correlation, which are potentially correct matches, as inputs to the next step; encoding of potential matches as second-order symmetric tensors.



**Figure 1. The Arena**

Tensor representation is capable of capturing a location's position along with the likelihood of its role as part of a smooth surface, a curve, or as an outlier. Geometrically, the second-order symmetric tensor is equivalent to an ellipsoid whose shape encodes the location's orientation information, and

whose size encodes the perceptual saliency, or confidence, of this information. Tensor voting [5] is the means for communicating information among neighboring locations in order to determine the amount of local support a feature receives. The votes are cast in the form of second-order symmetric tensors that can be accumulated at the receiving location by a simple addition of 3x3 matrices. The eigensystem of the resulting matrix contains information on the surface and curve saliency of the location along with the orientation of the corresponding elementary surface or

curve going through the location. If a location displays very low saliency, it can be discarded as an outlier.

Surface and curve extraction are performed using a modified Marching Cubes process [2]. Perceptual structures are extracted as the zero-crossings of the first derivative of surface or curve saliency with sub-voxel accuracy. Fig. 1 shows results from a binocular example.

3-D processing, implemented by the tensor voting framework, eliminates the problems associated with the 2 1/2-D sketch, which is view-centered, suffers at discontinuities, and cannot support multiple layers or transparency. Tensor voting enables us to perform all the necessary processing in three dimensions at a manageable computational cost.

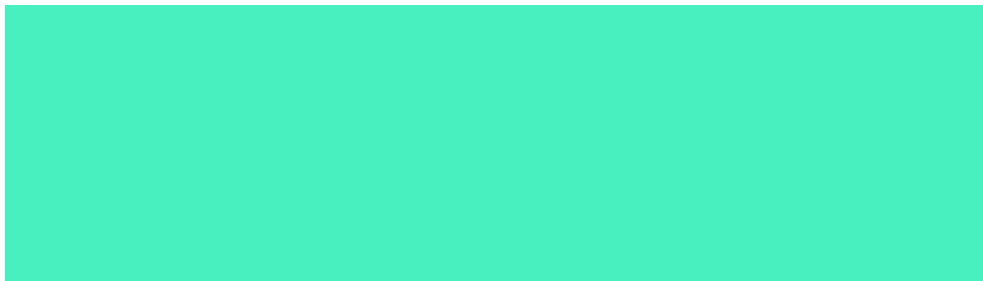
The novelty of our method is the fact that we delay decisions on the correctness of matches and do not make them at a premature stage. From our perspective, high cross-correlation values between image intensities are indications of potential matches, but not very reliable as a criterion for resolving the correctness of matches. We propose the use of saliency for that purpose. Saliency represents the likelihood that a location belongs to a perceptual structure, which could be a surface, a curve, or a junction. This enables us to integrate feature matching with surface and curve extraction. The matches, that are considered inliers, are used to guide the extraction process, and hence, their correctness can be judged by the surfaces they produce.

### **Extension to Multiple Views**

The design of the binocular system was made in a way that the system could be extended to multiple views with no modifications of its principal concepts. Three-dimensional processing is even more crucial in the case of multiple views, where the 2 1/2-D representation is inadequate. The fact that decisions for the correctness of matches are not made prematurely becomes more important, since there is more information stemming from more than two views that should be

utilized before such decisions can be made. We process images in pairs, under the principles of the binocular system, and increase the initial saliency of matches that are consistently supported by multiple pairs.

In order to maintain the general applicability of our method, we do not treat any images as privileged and do not require that features appear in all views. The only requirement for a feature to be used as input to the tensor voting process is that it appears in at least two views, so that an estimate of its depth can be obtained. The main additional requirement for processing multiple views is that correspondences have to be reconstructed in real world coordinates since disparity space is no longer sufficient. Fig. 2 (a) shows three of the 36 input images of the Dragon sequence, while Fig. 2 (b) shows some views of the output surfaces.



**Figure 2. Input images and output surfaces of the Dragon**

- [1] M.S. Lee and G. Medioni, *Inferring Segmented Surface Description from Stereo Data*, Proc. CVPR, pp. 346-352, 1998.
- [2] W.E. Lorensen and H.E. Cline, *Marching Cubes: A High Resolution 3-D Surface Reconstruction Algorithm*, Computer Graphics, vol. 21 (4), pp. 163-169, 1987.
- [3] D. Marr and T. Poggio, *A Theory of Human Stereo Vision*, Proc. Roy. Soc. London, vol. B204, pp. 301-328, 1979.
- [4] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Co., 1982.
- [5] G. Medioni, M.S. Lee, and C.K. Tang, *A Computational Framework for Segmentation and Grouping*, Elsevier Science, 2000.
- [6] A.L. Yuille and T. Poggio, *A Generalized Ordering Constraint for Stereo Correspondence*, AI Memo 777, AI Lab, MIT, 1984.