

An Experimental System for the Integration of Information from Stereo and Multiple Shape From Texture Algorithms

Terrance E. Boulton and Mark Moerdler
Columbia University Department of Computer Science
New York City, New York, 10027. tboulton@csc.columbia.edu mark@csc.columbia.edu

ABSTRACT

In numerous computer vision applications, there is both the need and the ability to access multiple types of information about the three dimensional aspects of objects or surfaces. When this information comes from different sources the combination becomes non-trivial.

This paper describes the present state of ongoing research in Columbia's Vision Laboratory in the integration of multiple visual sensing methodologies which yield three dimensional information, in particular, feature based stereo algorithms, and various shape-from-texture algorithms are already in operation and multi-view shape-from-texture and shape-from shading modules are expected to be incorporated. Unlike most systems for multi-sensor integration, which fuse all the information at one conceptual level, e.g., the surface level, the system under development uses two levels of data fusion, intra-process integration and inter-process integration. The paper discusses intra-process integration techniques for feature-based stereo and shape-from-texture algorithms. It also discusses a inter-process integration technique based on smooth models of surfaces. Examples are presented using camera acquired images.

1 INTRODUCTION

This paper discusses research, currently in progress in Columbia's Vision Laboratory, on the integration data from two different, but highly compatible modalities. The different data are the product of two feature based stereo algorithms which use different types of features and multiple shape-from-texture algorithms. These modalities are being considered because they are, in general, applicable to similar regions of an images. This allows experimentation with both corroborating information, e.g., stereo images of textured three dimensional surfaces, and conflicting information (not reported here), e.g., stereo images of a two dimensional image of a textured three dimensional surface.

While solving the general fusion problem is beyond the current abilities of AI research, there has been progress in restricted contexts; these including fusion of stereo and tactile data [Allen 85, Allen and Bajcsy 85], pointwise fusion of data [Henderson and Fai 83], and fusion of various information about an intensity image for the purpose of segmentation [Belknap et al. 85, Kohler 84, D. M. McKeown and McDermott 84]. There has also been work on the regularization as a means for fusion [Blake and Zisserman 86, Medioni and Yasumoto 85, Poggio and Torre 84].

All of the above mentioned system fuse their information at one conceptual level, e.g., the image or surface level. In contrast, the system under development uses two levels of data fusion, intra-process integration and inter-process integration. The former is fusion of information generated by all shape-from-X approaches with certain predetermined similarities, e.g., feature based stereo algorithms with different features. The latter type of integration is the fusion of the information resulting from each of the intra-process integration phases, with any *a priori* knowledge, e.g., smoothness assumptions or model assumptions. However, to allow for some amount of top-down processing, there is communication between each process through a global blackboard.

The techniques for intra-process integration are dependent on the assumptions, explicit and implicit, in the underlying processes. Brief presentation are given of two such intra-process integration techniques.

Inasmuch as the final result of the data fusion is assumed to be objects with smooth surfaces, the inter-process integration should depend on our model(s) of surfaces, not on the data acquisition techniques. The current system employs a regularization based surface reconstruction technique for this task. This approach allow the system to independently weight each piece of information from the intra-process integration phases. Part of this weighting is a global factor determining which of the modalities has higher priority.

The interaction among the various computational modules as well as the integration modules, is accomplished with a blackboard organization. This scheme allows bidirectional flows of information and provides a means for easy detection of when the information necessary for modules execution is present in the system. This portion of the system will not be considered further in this paper.

The remainder of this paper is divided into sections on background and motivation, texture algorithms and related intra-process integration, stereo algorithms and related intra-process integration, and inter-process integration. Following the description of the current system, the results of limited experimental testing, using camera images, is presented.

2 MOTIVATION AND BACKGROUND

The integration of sensory data from different modalities is an important and challenging problem for computer vision researchers.

As research in vision has progressed, researchers began to realize that the information available from a single "shape-from" algorithm would not be sufficient to solve the general vision problem. Prior vision research has yielded different "modalities" of information including:

numerous approaches to shape-from-texture [Kender 80, Witkin 80, Aloimonos 86], binocular stereo [Marr and Poggio 79, Pollard et al. 85], [Eastman and Waxman 85, Hoff and Ahuja 85], shape-from shading [Pentland 86, Horn 70, Ikeuchi 80a, Lee 85], and shape-from-motion [Anandan and Weiss 85, Prazdny 79]. Each of these sources of shape information has different domains of applicability, different computational complexity, and different error characteristics. For any given module, there would exist numerous images (or regions thereof) for which the module would not correctly predict surface shape. Some of the sources are complementary, e.g. shape-from-shading will apply generally only in those regions where shape-from-texture will fail. Other modules can act in either a competitive or synergistic fashion, e.g. binocular stereo and shape-from-texture will generally apply in the same regions of an image, and may compete for dominance if their outputs differ, or can they mutually reinforce a consistent interpretation.

There are two main reasons that this research has considered two separate levels of integration. The first is computational in nature, and derives from the fact that it is easier to heuristically combine data from similar sources, as in the intra-process integration phase. Then, when the system must integrate information from markedly different sources, that information should already be of higher quality than the initial raw data. This separation of duties also aids in maintaining system modularity, and minimizes global memory requirements.

The second reason for desiring a multi-level integration scheme follows from studies of human vision. Considerable research exists on the human perception of three-dimensional surfaces, e.g. see [Julesz 60, Uttal 83, Rogers and Graham 83, Bülthoff and Mallot 87]. Many of these works have used selective stimuli, e.g., random-dot stereograms, to study phenomenological aspects of depth perception from various sources. From these studies one can draw inferences about the integration of information from "modules" using this information. Others of these works have studied the "ordering" of operations in the human visual system and the interaction of various information sources for depth perception. Of particular relevance to the question of multi-level integration work along the lines of [Bülthoff and Mallot 87], which examined the interaction, both rivalrous and mutually supportive, among various information sources.

3 TEXTURE PROCESSES AND TEXTURE INTRA-PROCESS INTEGRATION

This section discusses our approach to the problem of deriving the orientation information from multiple independent textual cues. The generality of this approach is due to the interaction between textural cues, thus allowing it to extract shape information from a wider range of textured surfaces than any individual method. The method consists of two major phases: the calculation of orientation constraints and the generation of *texel patches*^{*}, and intra-process integration where constraints on each patch are fused into a "most likely" orientation.

The robustness of this approach has already been demonstrated elsewhere, see [Moerdler and Kender 87b, Moerdler and Kender 87a]. Currently the shape-from-texture methods used are: shape-from-uniform-texel-spacing [Moerdler and Kender 85], and shape-from-uniform-texel-size [Ohta et al. 81]. These two methods generate orientation constraints for different overlapping classes of textures.

3.1 Background

Current methods to derive shape-from-texture are based on measuring a distortion that occurs when a textured surface is viewed under perspective assuming of course that natural texture neither mimics nor cancels projective effects. The perspective distortion results in some aspect of the texture being deformed when the scene is imaged. In order to simplify the recovery of the orientation parameters from this distortion, researchers have imposed limitations on the applicable class of textured surfaces. Some of the limiting assumptions include uniform texel spacing [Kender 80, Kender 83, Moerdler and Kender 85], uniform texel size [Ikeuchi 80b, Ohta et al. 81], uniform texel density [Aloimonos 86], and texel isotropy [Witkin 80, Davis et al. 83, Dunn et al. 84]. These are strong limitations causing methods based on them to be applicable to only a limited range of real images.

3.2 Design Methodology

The generation of orientation constraints from perspective distortion is performed using one or more image texels. The orientation constraints can be considered as local, defining the orientation of individual surface patches called *texel patches*. Texel patches are defined by how each method utilizes the texels. Some methods, e.g., uniform texel size, use a measured change between two texels; in this case the texel patches are the texels themselves. Other methods, e.g., uniform texel density, use a change between two areas of the image. In the latter case the texel patches are predefined areas of the image. For the texture modules, intra-process fusion is carried out at the texel patch level. This differs from integration at the surface level which has been attempted elsewhere, e.g., [Ikeuchi 80b] and [Aloimonos 86] both develop integration algorithms that use constraint propagation and relaxation to derive a single orientation per surface patch.[†]

The process of fusing orientation constraints and generating surfaces can be broken down into the following three phases: (1) the creation of texel patches, (2) calculation of (multiple) orientation constraints for each texel patch, and (3) the unification of the orientation constraints per texel patch into a "most likely" orientation. Each of the remaining portions of this subsection describes one of these phases.

3.2.1 Texel patch definition

There has been considerable work in computer vision on the automatic recognition of textural patches. While accurate and consistent texel patch recovery would greatly simplify the integration process, we feel that such data is unavailable at the present time. Instead, we

^{*}A *texel patch* is a 2-D description of a sub-image that contains one or more textural elements. The number of elements that compose a patch is dependent on the shape-from-texture algorithm.

[†]These approaches were further limited by the use of only a single shape-from-texture method.

have chosen a simplistic patch definition obtained by first processing the image with assorted filters and then thresholding the image to define patches.² We acknowledge that better texture discrimination algorithms exist, but this was not the focus of our research. In the work described in this paper we have filtered the image by local averaging of the gray levels. We have also experimented with edge detection and edge orientation filters (with and without post filtering smoothing), although those filters are not used on the examples herein.

3.2.2 Surface Patch and Orientation Constraint Generation

The first phase of the system consisting of several shape-from-texture components which generate augmented texels. Each augmented texel consisting of a texel patch, orientation constraints for the texel patch, and a confidence weighting per constraint. The orientation constraints are stored in the augmented texel as vanishing points which are mathematically equivalent to a class of other orientation notations (e.g. pan and tilt constraints) [Shafer et al. 82]. Moreover, they are simple to generate and compact to store.

The confidence weighting is defined separately for each shape-from method and is based upon the intrinsic error of the method. For example, shape-from-uniform-texel-spacing's confidence weighting is a function of the total distance between the texel patches used to generate that constraint. A low confidence value is given when the inter-texel distance is small (1 texel distance) because under these conditions a small digitization error causes a large orientation error. Above the threshold the confidence weighting is set high (6) and then starts to decrease as the inter-texel distance increases. Mathematically, given the radius of the texel, *radius*, and the inter-texel distance, *distance*, the confidence weighting *W* is: $W = 6 \times \frac{radius}{distance}$.

The decreasing confidence occurs because once the inter-texel distance grows too large the local surface is no longer approximated by a plane and the orientation error grows. This further acts to make the constraints group locally rather than globally which is valid since texels that are part of the same surface are normally located close together.

The current system contains two shape-from-texture methods, shape-from-uniform-texel-spacing [Moerdler and Kender 85], and shape-from-uniform-texel-size [Ohta et al. 81]. Each of the methods is based on a different textural characteristic that allows the generation of orientation constraints and also limits the applicability of the approach. Future plans call for the inclusion of shape-from-texture isotropy ala [Witkin 80], and shape-from-ellipticity of circular textures.

Shape-form-uniform-texel-spacing derives orientation constraints based on the assumption that the texels can be of arbitrary shape but are equally spaced, while shape-from-uniform-texel-size is based on the unrelated criteria that the spacing between texels can be arbitrary and the size of all of the texels is equivalent but unknown.

In shape-from-uniform-texel-size if the distance from the center of mass of texel T^1 to texel T^2 is defined as D then the distance from the center of texel T^2 to a point on the vanishing line can be written as :

$$F_1 = D \times \frac{S_3^{\dagger}}{S_1^{\dagger} - S_2^{\dagger}} \quad (1)$$

In shape-from-uniform-texel-spacing the calculations are similar. Given any two texels T^1 and T^2 inter-texel distance is defined as D , if the distance from T^1 to a mid-texel T^3 is equal to L and the distance from T^2 to the same mid-texel T^3 is equal to R , the distance from texel T^1 to a vanishing point is given by :

$$F_2 = \frac{D \times (R + 1)}{L - R} \quad (2)$$

3.3 Intra-Process Integration for Textures: Generation of most likely orientation

Once the orientation constraints have been generated for each augmented texel, the next step consists of unifying the constraints into one orientation per augmented texel. The major difficulty in deriving this "most likely" orientation is that the constraints are errorful, inconsistent, and potentially incorrect. A simple and computationally feasible, solution to this is to use a Gaussian Sphere which maps the orientation constraints to points on the sphere [Shafer et al. 32]. A single vanishing point circumscribes a great circle on the Gaussian Sphere; two different constraints generate two great circles that overlap at two points uniquely defining the orientation of both the visible and invisible sides of the surface patch.

The Gaussian sphere is approximated within the module by the hierarchical tessellated Gaussian Sphere based on triangular shaped faces called trixels [Fekete and Davis 84, Korn and Dyer 86]. The top level of the hierarchy is the twenty face icosahedron. At each level, other than the lowest level of the hierarchy, each trixel has four children which more closely approximate the curvature of the spherical surface than their parent. This hierarchical methodology allows the user to specify the accuracy to which the orientation can be calculated by defining the number of levels of tessellation that are created.

The intra-texture-process integration phase generates the "most likely" orientation for each texel patch by accumulating the evidence from all the orientation constraints (generated in phase one) for the patch. For each constraint, it initially visits the twenty top level trixels, determines whether the great circle falls on the trixel and if the result is positive, visits the children. At each lowest level trixel through which the great circle travels, the likelihood value of the trixel is incremented by the constraint's weight. The hierarchical nature of this approach limits the number of trixels that need to be visited. Once all of the constraints for a texel patch have been considered, a peak

²An analysis of this type of technique for texture discrimination can be found in [Davis et al. 84].

finding program smears the likelihood values at the lowest level trixels. The "most likely" orientation is defined to be the trixel with the largest smeared value.

This method does not assure that under all circumstances a single "most likely" orientation will be derived. When more than one "most likely" orientation is derived for a patch the module performs a Waltz type filtering. It computes the "most likely" orientation for the remaining augmented texels and then re-analyzes the orientation constraints for each texel that does not have a single "most likely" orientation. For each unsolved texel patch the module considers all of the patch's constraints and removes any constraints that do not correctly define the "most likely" orientation of another texel patch. Once this constraint pruning has occurred the module recomputes the "most likely" orientation for the patch. This secondary analysis does not assure a single "most likely" orientation either, but it does aid in simplifying and deriving a single "most likely" orientation for the largest number of surface patches.

Under certain conditions either method may generate incorrect constraints which the intra-texture-process integration phase has effectively ignored. On textures that are solvable by both methods the methods cooperate and correctly define the textured surface(s) in the image.

4 STEREO PROCESSES AND INTRA-PROCESS INTEGRATION

The stereo-based processes of the system are based on feature matching between the two images. The system uses multiple feature definitions to insure both good localization and noise resistance. These features are then classified as to amount of ambiguity. The system starts with the least ambiguous matches and reconstructs a disparity surface. Intra-stereo-integration is accomplished through a regularized reconstruction of the disparity field based on the supposition that the smooth surfaces in the world give rise to a smooth disparity surface. After all points are considered, the intra-process module adds its output to the blackboard. Currently this output is depth values at various points, especially along the "edges" of surfaces in the disparity field and at the locations of feature points.

4.1 Definition of the multiple features

A common problem in stereo systems is that the features are too sparse, have poor localization, or are sensitive to noise. Rather than attempt to define yet another feature for matching, the stereo module currently combines two different types of features. These are: (1) zero crossings of laplacian of gaussians of the images, which are subsequently thresholded (based on magnitude of crossing) and matched along approximately epi-polar lines using orientation and sign as filters, and (2) centroids of texels defined in the shape-from-texture algorithm (with some of the other texel features used to insure only valid matches). The first of these features provide a large number of features for the matching algorithm, unfortunately the localization of these features are not highly accurate. The second set of features are not very dense, however, they provide very accurate localization of the feature.

In the future we will be adding features derived from area based correlations, interest operators (e.g. [Moravec 79]) and a thresholded sobel operator.

4.2 Intra-stereo-integration module

The integration of the various features is accomplished by a multi-pass matching algorithm, where the quality of localization/ambiguity effects the order in which points are considered, and previously matched points effect the disambiguation of other points. The matching algorithm used is described in detail elsewhere, [Boult and Chen 87], only a brief description is presented here. The basic assumption underlying the most matching algorithm is that the disparity surface should be smooth. In vision research, there have been many matching models algorithms proposed with different "smoothly" varying disparities, [Marr and Poggio 79, Mayhew and Frisby 81], [Eastman and Waxman 85]. The smooth disparity fields used for this system are based on generalized two dimensional smoothing splines, see [Boult 86]. The smoothness criterion is similar to one used in smooth surface reconstruction, see [Blake and Zisserman 86, Boult 86, Grimson 79, Terzopoulos 84].

The system starts with the feature points which have "unique matches" and good localization (i.e., at the current time it begins with centroids of "texel" defined on the blackboard). In all neighborhoods without these features, lower quality (in terms of localization) features with "unique" matches are added, however they are given a lower confidence value. Thus when the smooth surface is fitted to the disparity data, the disparity values generated by lower quality features will not be as closely approximated.

After all the "unique" matches have been used, the module reconstructs a disparity surface. This reconstruction is based on the assumption that the disparity surface should give rise to a smooth surface in depth. Using this disparity surface, the module disambiguates other matches by choosing the potential match which comes closest to the smooth surface. The distance between the disparity predicted by the "best" match and the smoothed disparity surface affects the confidence of the match, which in turns affects the way the disparity surface approximates that match. The disambiguation takes place in multiple passes each of which incorporates to features that are increasingly ambiguous.

After all points are considered, the intra-process module adds its output to the blackboard. Currently, this output is depth values at various points, especially along the "edges" of surfaces in the disparity field, and depends on the calibration of the imaging system, which is based on [Tsai 86]. Future plans call for the module to output surface orientation information in the place of depth data, thereby eliminating the need for calibration, see [Boult and Chen 87]

5 INTER-PROCESS INTEGRATION AND SURFACE RECONSTRUCTION

This section describes the inter-process integration phase of the system. This phase of the fusion process is predicated on the assumption that the world is comprised of piecewise smooth surfaces/objects. Therefore, the inter-process integration should depend on the assumed

smoothness model(s) for surfaces, not on the data acquisition techniques. There are two main aspects of the inter-process integration, basic surface building, and the weighting of various modules.

5.1 Basic Surface Reconstruction Technique

Inasmuch as they can be expressed in terms of inverse optics, many problems in computer vision, including surface reconstruction from sparse 3D information, are ill-posed. One way of reformulating these ill-posed problems is through a well known technique called regularization.^{**} Let us precisely define the problem at hand:

Let $\mathcal{F}_1$, the space of allowed surfaces, be a Hilbert or semi-Hilbert space. Let $\|\cdot\|_{\mathcal{F}_1}$ be a norm (or semi-norm if $\mathcal{F}_1$ is semi-Hilbert) measuring the "unreasonableness" of a surface f . Let $N_n(f) = [L_1(f), \dots, L_n(f)]$ be the given information. Then the visual surface reconstruction problem is to find $f^* \in \mathcal{F}_1$ such that

$$\vartheta(f) = \min_{f \in \mathcal{F}_1} \vartheta(\tilde{f}) \text{ where } \vartheta(\tilde{f}) \stackrel{\text{def}}{=} \lambda \cdot \|\tilde{f}\|_{\mathcal{F}_1} + \sum_{i=1}^n \left| \delta_i \times L_i(\tilde{f}) - L_i(f) \right|^2$$

The norm (semi-norm) $\|\cdot\|_{\mathcal{F}_1}$ is generally referred to as the *stabilizing functional* of the regularization. The class of functions $\mathcal{F}_1$ is an often overlooked, but immensely important, part of the regularization. One cannot indiscriminately choose how to regularize a problem. As pointed out in [Poggio et al. 85, page 315]^{††},

"... standard regularization methods have to be applied after a careful analysis of the ill-posed nature of the problem. The choice of the norm $\|\cdot\|_{\mathcal{F}_1}$ of the stabilizing functional $\|\cdot\|_{\mathcal{F}_1}$ and of the functional spaces involved is dictated by both mathematical properties and by physical plausibility. They determine whether the precise conditions for a correct regularization hold for any specific case."

The use of regularization for reconstruction of smooth surfaces in vision was first proposed in [Grimson 79]. In that pioneering work Grimson discussed the choice of the most appropriate stabilizing functional (though in different terminology). However, his decision was partially based on an assumed relationship between the stabilizing functional and the zero-crossings of the intensity images which lead to the stereo depth data. With respect to visual surface reconstructions, many researchers (e.g., [Terzopoulos 84, Hoff and Ahuja 85, Poggio et al. 85], [Lee 85, Choi and Kender 85]) seem to have accepted the choice of norm, stabilizing functional and associated functional spaces that were initially proposed in [Grimson 79].^{‡‡}

However, the authors feel that this class is "too smooth" and thus have adopted the assumption that world surfaces can be piecewise modeled as surfaces belonging to the class of $D^{-2}H^{-.5}$ with the second Sobolev semi-norm, which can intuitively be described as having only 1.5 derivatives in L^2 .^{***} In work reported elsewhere, [Boult 87], this assumption was shown to be at least as reasonable as assuming surfaces in $D^{-2}L^2$. The reconstruction of the surfaces might be accomplished with discrete regularization techniques. However, the approach taken herein is based on generalized smoothing spline functions, see [Boult 86, Franke 82, Wahba 84, Bates and Wahba 82, Meinguet 79], and is more efficient for sparse data, see [Boult 85, Franke 79].

The system allows for each data point to be individually weighted in the contribution to the allowed fitting error (the terms δ_i in the above definition) the choice of these weights is influenced by two things, the confidence passed for the point from the intra-process integration processes, and the weighting assigned to the module as a whole.

5.2 Weighting the outputs of the various intra-process integration modules.

The above fusion scheme requires that each data point be given a weight. The correct selection of these weights is difficult. The study of these weighting will be of paramount importance in our future research. Currently, the system builds three surfaces

1. one surface from the output of the intra-texture-integration module using the weighting supplied by that module.
2. one surface from the output of the intra-stereo-integration module using the weighting supplied by that module, and
3. one surface combining all data. For the combination, the weights are divided by the number of data-points output by a module. This provides some means for the texture data to have an effect on the surface. Otherwise the stereo data (with 500-5000 points) would totally dominate the shape information from texture (which only provides 10-50 data points).

While it would be nice for the system to choose which of these surface is the "correct" one, this is not possible. When the information is conflicting, the "correct" precept is subjective, and can often be changed by will in humans. However, when the surfaces agree, the system should be able to (but currently cannot) take note, and remove the redundant representations.

6 EXPERIMENTATION

The system described in this paper is still under development and has only been subject to limited experimental testing. Presented in this section are two examples of the system working on camera images. The results are presented in figures 1 to 6. One example, the curved

^{**}There exist numerous ways to calculate such surfaces, see [Boult 86, Chapter 9] for a critical comparison of 4 methods.

^{††}The mathematical notation in the quote has been modified to match that used in this paper.

^{‡‡}It is not clear if this acceptance is simply because the method gave reasonable results and there was a lack of alternatives, or because the choice is actually the most appropriate.

^{***}For a precise definition see [Boult 86, Boult 87, Meinguet 79].

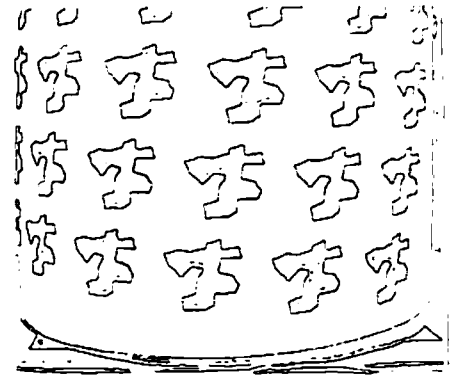
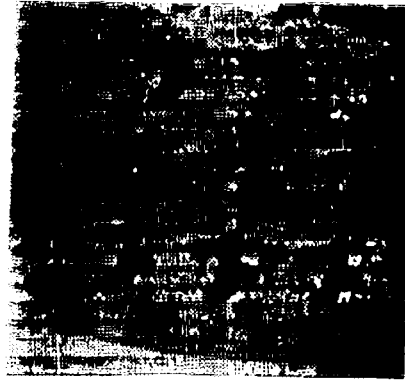
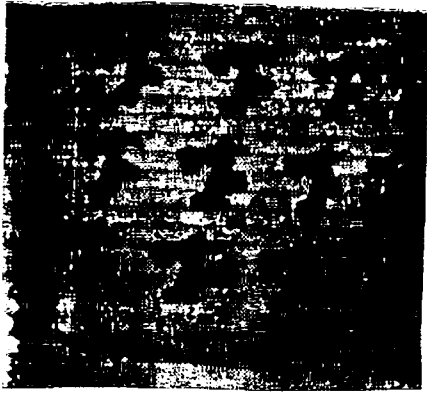


Figure 1: Left input image for example 1: an artificially textured roll of paper.

Figure 2: Right input image for example 1.

Figure 3: Laplacian of Gaussian of left image for stereo matching.

roll of paper demonstrates a surface where stereo dominates, but is significantly aided by the texture information. In the other example, texture was more successful because the scan line coherence assumed in stereo was violated by a slight rotation of the scene. The stereo module was thus slightly off on each "texel" and produced more of a bumpy surface (not shown).

7 CONCLUSIONS AND FUTURE PLANS

This paper describes an ongoing research project in Columbia's vision laboratory. The system discussed integrates two modalities: stereo and texture. The system makes use of a two level integration scheme, and while the experimental analysis is not complete, initial results show this multi-level integration to be both efficient and effective.

Future work will include the addition of modules for shape-from-multi-view-texture, shape-from-shading, and possibly shape-from-motion. In addition, the number of processes in both the existing stereo and texture modules will be expanded.

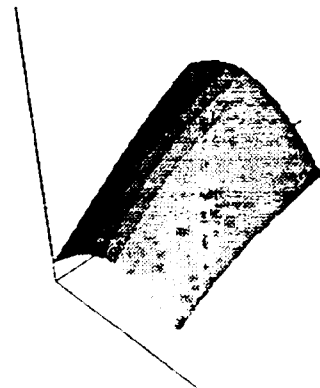
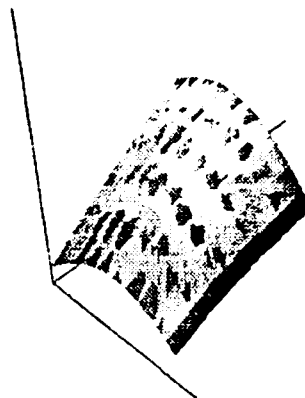
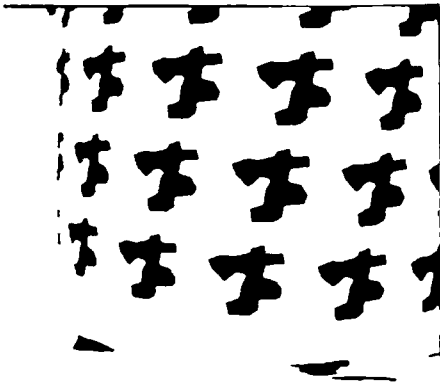


Figure 4: Texels identified in left image.

Figure 5: Reconstruction from just stereo data.

Figure 6: Reconstruction from combined texture and stereo outputs.

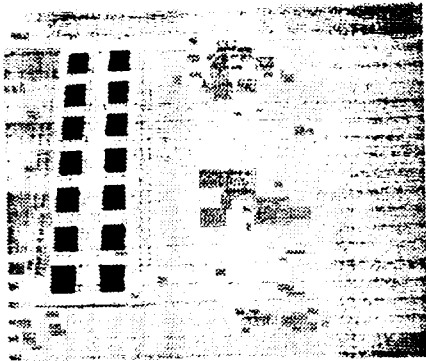


Figure 7: Left input image for example 2 a "circuit breadboard".

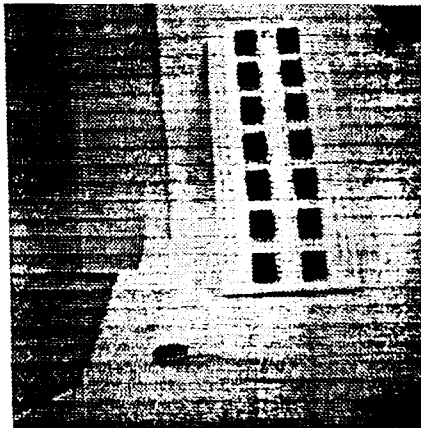


Figure 8: Right input image for example 1.

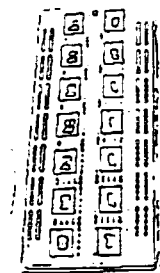


Figure 9: Laplacian of Gaussian of left image for stereo.

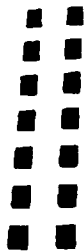


Figure 10: Texels identified in left image.

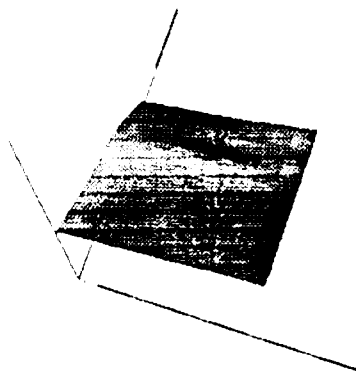


Figure 11: Reconstruction from only texture data.

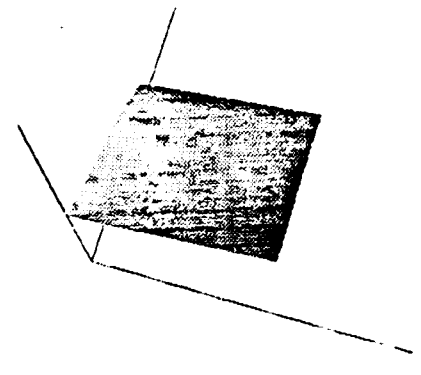


Figure 12: Reconstruction from combined texture and stereo outputs.

An important task awaiting these researchers is the study of the weighting of the outputs of the various modules, and some attempt at development of non-ad hoc criterion.

8 ACKNOWLEDGMENTS

This work was supported in part by Darpa Contract #N00039-84-C-0165. The authors are grateful to John Kender for his many discussions on this subject. Also thanks to L.H. Chen without whom the stereo modules would not have been finished this year.

REFERENCES

- [Allen 85] P. K. Allen. *Object Recognition Using Vision and Touch*. PhD thesis. University of Pennsylvania, Department of Computer Science., 1985.
- [Allen and Bajcsy 85] P. Allen and R. Bajcsy. Integrating sensory data for object recognition tasks. In *SPIE Image Processing Symposium*, Cannes, France, December 1985.
- [Aloimonos 86] J. Y. Aloimonos. Detection of surface orientation from texture, i: the case of planes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 584-593, June 1986.
- [Anandan and Weiss 85] P. Anandan and R. Weiss. Introducing a smoothness constraint in a matching approach for the computation of displacement fields. In *Proceedings of the DARPA Image Understanding Workshop*, pages 186-195, DARPA, 1985.
- [Bates and Wahba 82] D. Bates and G. Wahba. Computational methods for generalized cross-validation with large data sets. In C.T.H. Baker and G.F. Miller, editors, *Treatment of Integral Equations by Numerical Methods*, pages 283-296. Academic Press, New York, 1982.
- [Belknap et al. 85] R. Belknap, E. Riseman, and A. Hanson. The information fusion problem and rule-based hypotheses applied to complex aggregation of image events. In *Proceedings of the DARPA Image Understanding Workshop*, pages 279-292, DARPA, 1985.
- [Blake and Zisserman 86] A. Blake and A. Zisserman. Invariant surface reconstruction using weak continuity constraints. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 62-68, IEEE, 1986.
- [Boult 85] T. E. Boult. Visual surface interpolation: a comparison of two methods. In *Proceedings of the DARPA Image Understanding Workshop*, pages 446-478, DARPA, 1985.
- [Boult 86] Terrance E. Boult. *Information Based Complexity in Non-Linear Equations and Computer Vision*. PhD thesis. Department of Computer Science, Columbia University, 1986.
- [Boult 87] T.E. Boult. What is regular in regularization? In *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pages 457-462, IEEE, June 1987.
- [Boult and Chen 87] T.E. Boult and L.H. Chen. Analysis of two new stereo matching algorithms. *paper in preparation.*, 1987.
- [Bülthoff and Mallot 87] H.H. Bülthoff and H.A. Mallot. Interaction of different modules in depth perception. In *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pages 295-305, IEEE, June 1987.
- [Choi and Kender 85] D.J. Choi and J. R. Kender. Solving the depth interpolation problem with adaptive chebyshev acceleration method on a parallel computer. In *Proceedings of the DARPA Image Understanding Workshop*, pages 219-223, DARPA, 1985.
- [Davis et al. 83] L. S. Davis, L. Janos, and S.M. Dunn. Efficient recovery of shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(5):485-492, September 1983.
- [Davis et al. 84] L. S. Davis, A. Rosenfeld, and J.S. Weszka. *Region Extraction by Averaging and Thresholding*. Technical Report TR-311, F44620-72C-0062, University of Maryland Center For Automation Research, JUNE 1984.
- [D. M. Mckeown and McDermott 84] W.A. Harvey D. M. Mckeown and J. McDermott. *Rule-based Interpretation of Aerial Imagery*. Technical Report, CMU, September 1984.
- [Dunn et al. 84] S. M. Dunn, L. S. Davis, and H.A. Hakalahti. *Experiments in Recovering Surface Orientation from Texture*. Technical Report CAR-TR-61, University of Maryland Center For Automation Research, May 1984.
- [Eastman and Waxman 85] R.D. Eastman and A.M. Waxman. Disparity functionals and stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 245-254, DARPA, 1985.
- [Fekete and Davis 84] G. Fekete and L. S. Davis. Property spheres: a new representation for 3-d object recognition. In *Proceedings of the IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 192 - 201, IEEE, 1984.
- [Franke 79] R. Franke. *A Critical Comparison of Some Methods for Interpolation of Scattered Data*. Technical Report, Naval Postgraduate School, 1979. TR#NPS-53-79-003, Available from NTIS. #AD-A081 688/4.
- [Franke 82] R. Franke. Smooth interpolation of scattered data by local thin plate splines. *Comp. & Math. with Applications*, 8(4):273-281, 1982.
- [Grimson 79] W. E. L. Grimson. *From Images to Surfaces: A Computational Study of the Human Visual System*. PhD thesis, MIT, 1979.
- [Henderson and Fai 83] T. Henderson and W.S. Fai. *Pattern Recognition in a Multi-Sensor Environment*. Technical Report 83-001, University of Utah, Salt Lake City, July 1983. UUCS technical report.
- [Hoff and Ahuja 85] W. Hoff and N. Ahuja. Surfaces from stereo. In *Proceedings of the DARPA Image Understanding Workshop*, pages 98-106, DARPA, 1985.
- [Horn 70] B.K.P. Horn. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. *Technical Report MAC-TR-79, ()*, Project MAC, MIT 1970.
- [Ikeuchi 80a] K. Ikeuchi. *Numerical Shape from Shading and Occluding Contours in a Single View*. AI Lab Memo 566, MIT Artificial Intelligence Lab, February 1980.

- [Ikeuchi 80b] K. Ikeuchi. *Shape from Regular Patterns (An Example of Constraint Propagation in Vision)*. AI Lab Memo 567. MIT Artificial Intelligence Lab, March 1980.
- [Julesz 60] B. Julesz. Binocular depth perception of computer generated patterns. *Bell Systems Technical Journal*, 39:1125-1161, 1960.
- [Kender 80] J.R. Kender. *Shape from Texture*. PhD thesis, CMU. November 1980. Accepted for publication by Pitman Press, 1984.
- [Kender 83] J.R. Kender. Environmental labelings in low-level image understanding. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, page , August 1983.
- [Koenderink and vanDoorn 76] J. J. Koenderink and A. J. van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29-35, 1976.
- [Kohler 84] R. R. Kohler. *Integrating Non-Semantic Knowledge into Image Segmentation Processes*. COINS Technical Report 04. U. Mass. at Amherst, March 1984.
- [Korn and Dyer 86] M. R. Korn and C. R. Dyer. *3-D Multiview Object Representation for Model-Based Object Recognition*. Technical Report RC 11760, IBM T.J. Watson Research Center, March 1986.
- [Lee 85] D. Lee. *Contributions to Information-based Complexity, Image Understanding, and Logic Circuit Design*. PhD thesis, Department of Computer Science, Columbia University, 1985.
- [Marr and Poggio 79] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceeding Royal Society of London..* B(204):301-328, 1979.
- [Mayhew and Frisby 81] J.E.W. Mayhew and J. P. Frisby. Psychological and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349-385, 1981.
- [Medioni and Yasumoto 85] G. Medioni and Y. Yasumoto. Robust estimation of 3-d motion parameters from a sequence of image frames using regularization. In *Proceedings of the DARPA Image Understanding Workshop*, pages 117-128, DARPA, 1985.
- [Meinguet 79] J. Meinguet. Multivariate interpolation at arbitrary points made simple. *J. of Applied Mathematics and Physics (ZAMP)*, 30:292-304, 1979.
- [Moerdler and Kender 85] M. L. Moerdler and J. R. Kender. *Surface Orientation and Segmentation from Perspective Views of Parallel-Line Textures*. Technical Report, Columbia University, 1985.
- [Moerdler and Kender 87a] M. L. Moerdler and J. R. Kender. An approach to the fusion of multiple shape from texture algorithms. In *Proceedings of the Workshop on Spatial Reasoning and Multi-Sensor Fusion*, Morgan Kaufman Publishers, Inc., 1987.
- [Moerdler and Kender 87b] M. L. Moerdler and J. R. Kender. An integrated system that unifies multiple shape from texture algorithms. In *Proceedings of AAAI 87*, American Association for Artificial Intelligence, Morgan Kaufman Publishers, Inc., 1987.
- [Moravec 79] H.P. Moravec. Visual mapping by a robot rover. In *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, pages 598-600, August 1979.
- [Ohta et al. 81] Y. Ohta, K. Maenobu, and T. Sakai. Obtaining surface orientation from texels under perspective projection. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, IJCAI, IJCAI*, 1981.
- [Pentland 86] Alex P. Pentland. Shading into texture. *Artificial Intelligence*, 29(2):147-170, August 1986.
- [Poggio and Torre 84] T. Poggio and V. Torre. *Ill-posed Problems and Regularization in Early Vision*. Technical Report lab memo 773, Massachusetts Institute Technology AI, 1984. Also appeared in the Proceedings of the DARPA Image Understanding Workshop.
- [Poggio et al. 85] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(6035):314-319, 1985.
- [Pollard et al. 85] S.B Pollard, J.E.W. Mayhew, and J.P. Frisby. Pmf: a stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449-470, 1985.
- [Prazdny 79] K. Prazdny. *Egomotion and Relative Depth Map from Optical Flow*. PhD thesis, Computer Science Department, University of Essex, 1979.
- [Rogers and Graham 83] B. J. Rogers and M. E. Graham. Anisotropies in the preception of three-dimensional surfaces. *Science*, 221:1409-1411, 1983.
- [Shafer et al. 82] S.A. Shafer, T. Kanade, and J.R. Kender. Gradient space under orthography and perspective. In *Proceedings of the IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 26-34, August 1982.
- [Terzopoulos 84] D. Terzopoulos. *Multiresolution Computation of Visible-Surface Representations*. PhD thesis, MIT, 1984.
- [Tsai 86] R.Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 364-374, June 1986.
- [Uttal 83] W. R. Uttal. *Visual Form Detection in 3-Dimensional Space*. Lawrence Erlbaum Associates, Hillsdale NJ, 1983.
- [Wahba 84] G. Wahba. Surface fitting with scattered noisy data on euclidean d-space and on the sphere. *Rocky Mountain Journal of Mathematics*, 14(1):281-299, 1984.
- [Witkin 80] A. Witkin. Recovering surface shape from orientation and texture. In M. Brady, editor, *Computer Vision*, North-Holland, 1980.