Binocular and Multiple View Stereo Using Tensor Voting

1. Research Team

Project Leader: Prof. Gérard Medioni, *Computer Science*

Graduate Students: Philippos Mordohai

Industrial Partner(s): Geometrix Inc

2. Statement of Project Goals

We propose to treat the problem of binocular and multiple view stereo from a perceptual organization perspective using the Tensor Voting framework we have developed [4]. Under our approach feature matching and structure reconstruction are addressed within the same framework. In order to handle noise, lack of image features, and discontinuities, we adopt a tensor representation for the data and tensor voting for information propagation.

3. Project Role in Support of IMSC Strategic Plan

Our ability to generate models from images can be very useful to other projects at IMSC. It provides models to research related to graphics, animation, and haptics without having to resort to expensive and cumbersome laser range scanners. Other possible applications of our research in conjunction with other IMSC research projects include tele-conferencing and immersipresence, e-commerce, scene analysis and cartography.

4. Discussion of Methodology Used

One aspect of this research is binocular stereo. Literature on the subject is extensive [8]. A predecessor of our current research, where stereo is addressed in a perceptual organization framework, was published in [1,3]. Over the last year we focused on arguably the largest remaining difficulty in binocular stereo, which relates to occlusion. Occluded pixels are a major source of erroneous matches, since they can achieve high matching scores. We propose a binocular stereo algorithm within the Tensor Voting framework, where monocular information from the images is used to resolve problems caused by occlusion. Initial matching is performed by multiple-window cross-correlation and points are grouped into surfaces by tensor voting. We show on real stereo pairs with ground truth that we are not only able to estimate the depth maps, but that we can also get reliable depth estimates for the occluded pixels.

The most significant contribution in 2002 was the development of a module that detects occlusion boundaries on the images using tensor voting in 2-D. The boundaries of the segmented surfaces from the stereo algorithm serve as hypotheses for the location and orientation of the actual occlusion boundaries. A robust method for extracting these boundaries combining the prior information and image edges has been submitted for review to an international conference. Once the correct occlusion edges have been detected, the intermediate extracted surfaces are adjusted accordingly. Some results on standard stereo pairs can be seen in Figure 1.

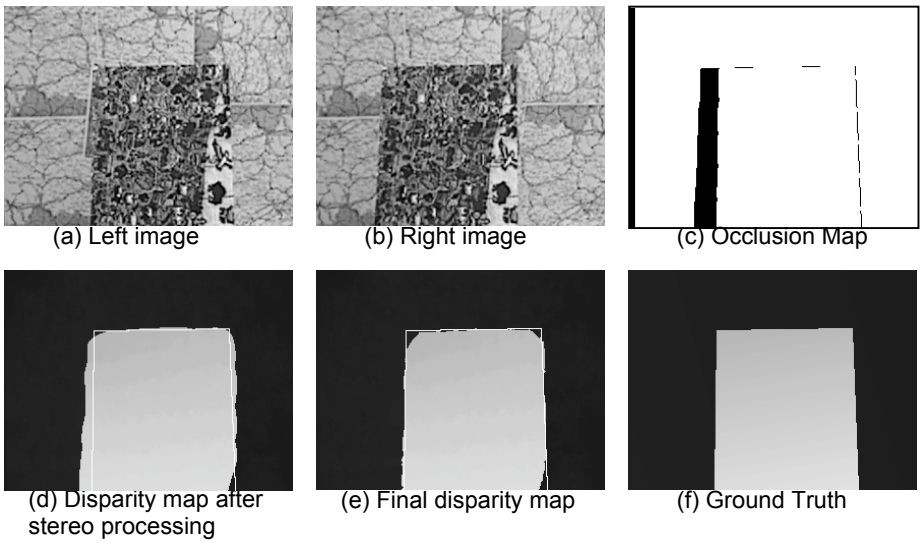

Figure 1. Results on the "map" stereo pair

On the other hand, we also address the problem of multiple view stereo as a perceptual organization problem using Tensor Voting. The framework enables us to perform all processing in three dimensions after combining the initial information from all views. The computational cost is manageable, since voting is performed locally and complexity is linear with respect to the number of views. We also presented results on real image sets, such as the ones seen in Figure 2.

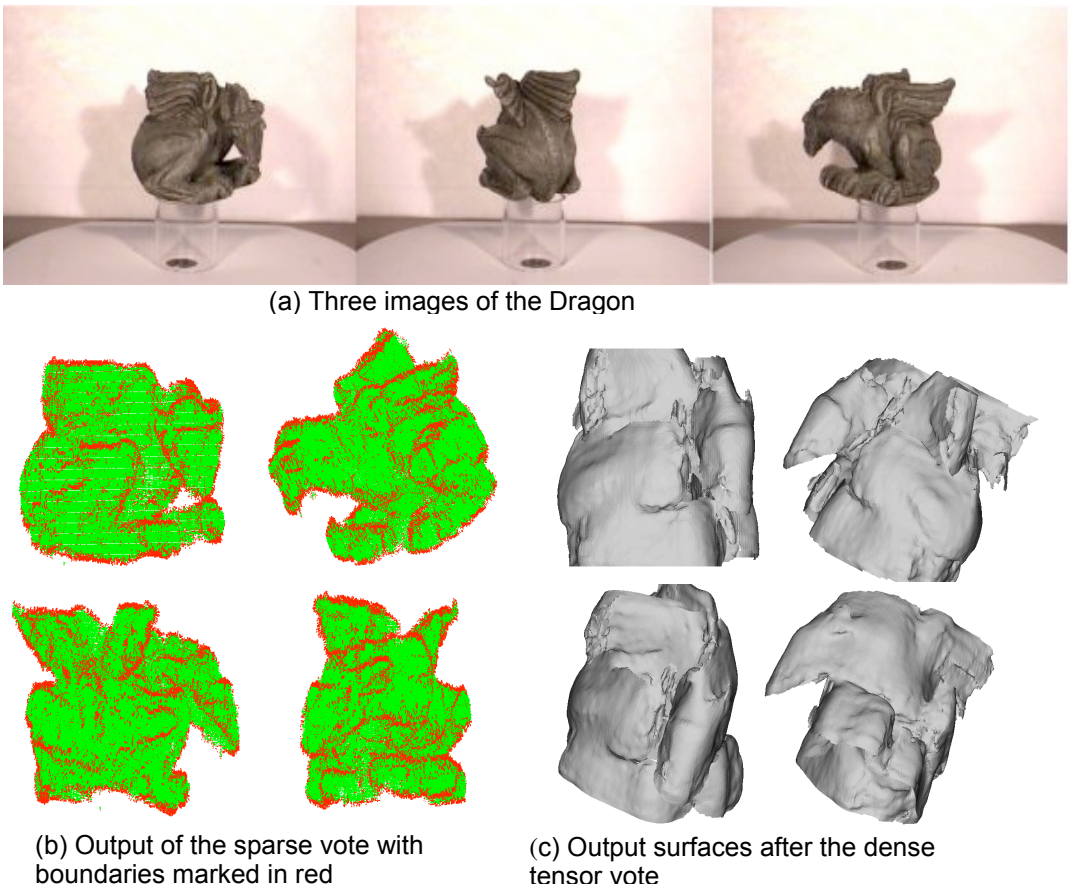

Figure 2. The Dragon Sequence

5. Short Description of Achievements in Previous Years

During the initial phase of this research, the Tensor Voting framework was developed and tested. Techniques for extracting regions, surfaces and curves from noisy three-dimensional datasets were developed [2] and stereovision was an excellent area to apply them. The problem of stereo reconstruction is a challenging and demanding one, thus aiding the development of our methodology, while, on the other hand, the Tensor Voting framework is well suited for the various problems that arise in the processing of stereo data. After our original binocular stereo approach was published [1,3], we focused our attention on achieving performance comparable to the state-of-the-art in the binocular case and on developing a leading approach to multiple view stereo.

In 2001, we further confirmed the validity of our approach by applying it to classical stereo pairs and achieving good results. We have enforced constraints that are specific to the problem at hand via the Tensor Voting framework. In the case of multiple view stereo, the concept of consistency of matches was introduced by increasing the initial saliency of matches confirmed by more than two views. These modifications proved very effective as demonstrated by the improvement in the output [6].

As far as the Tensor Voting framework is concerned, there were significant steps of progress, the most effective of which so far was the introduction of first-order tensor information that complements the second-order information and accommodates very accurate discontinuity and boundary detection. Boundaries play a major role in perceptual organization and data visualization, and they are the foundation of the multiple-scale approach we plan to use. We also plan to conduct extensive experiments on multi-scale processing of data and on the relationships and transformations between the different scales.

5a. Detail of Accomplishments During the Past Year

In 2002, our preliminary approach to binocular stereo was published in the most highly esteemed journal in the field, the IEEE Transactions on Pattern Analysis and Machine Intelligence [3]. Later, a paper on multiple view stereo was presented in the International Conference of Pattern Recognition sponsored by the IEEE and the IAPR [6]. Finally, we invited to contribute a chapter, containing both details of the Tensor Voting framework and results on stereo, to a book titled "Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics".

We were able to achieve very encouraging results on evaluation datasets for binocular stereo, proposed by D. Scharstein and R. Szeliski [8]. In fact, the performance of our algorithm would rank it eighth among 22 algorithms that have submitted results at the Middlebury College Stereo Research Evaluation Page (http://www.middlebury.edu/stereo/). These results were submitted for review to the International Conference on Computer Vision and Pattern Recognition (CVPR 2003).

6. Other Relevant Work Being Conducted and How this Project is Different

Due to lack of space we refer interested readers to [8] for a comprehensive review of binocular stereo algorithms. A multitude of approaches have been used including dynamic programming, non-linear optimization, relaxation labeling, clustering, graph cuts, stochastic optimization, non-linear diffusion, cooperative stereo and adaptive windows. Each method has different advantages and can be applied to broader or more limited types of scenes according to the assumptions made. Good results have been obtained by all the afore-mentioned techniques.

As far as multiple view stereo is concerned, volumetric methods have gained increased popularity and have achieved very good results. Surveys of these methods were published in [9]. Volumetric methods adopt a ray-space representation of the scene and pose multiple view stereo as the computation of the visual hull of the scene. Even though they are currently the state of the art, they suffer from some inherent limitations. First and foremost among them is the inability to handle concavities in the scene, while the general placement of cameras further increases the computational cost [9]. Volumetric methods are by nature iterative, require careful initialization of the scene model, and cannot recover from erroneous initializations. Perceptual methods operate at a higher level than pixels and voxels exploiting relationships between elements of the same scene or image structure. A perceptual grouping framework enables the propagation of information among tokens and imposes global constraints, such as smoothness; therefore, it can utilize information from the images more efficiently. The limitations of current approaches based on perceptual organization are that either their representation is inadequate for generalization to multiple views, or that computational complexity increases exponentially with the number of views.

7. Plan for the Next Year

During the next year we intend to investigate further the use of monocular, image-based cues in our stereo algorithm. They provide information that we had not fully utilized so far, as we developed the 3-D part of the algorithm. As soon as a stable and efficient binocular version is on hand, we intend to extend this knowledge to the multiple view domain.

We also intend to investigate the application of Tensor Voting in multiple scales, as many real scenes contain objects that are perceived at different scales. This is due to variations in size and level of detail as well as variations in texture density. A multiple scale processing scheme is necessary for processing such scenes. It must be capable of detecting and preserving fine details, not allowing larger or richer in texture regions to dominate smaller or less textures ones, achieving good continuation of structures, and filling in missing data. Instead of compromising the quality of the results by selecting a single scale of analysis for the entire scene, which would be sub-optimal in most regions, we intend to develop an approach that automatically infers the smallest scale that can preserve the finer details, then proceeds with larger scales to ensure good continuation in regions with sparse or missing data.

8. Expected Milestones and Deliverables

We intend to submit our methodology and results to conferences and journals for review and publication. The source code and our results on stereo will be submitted to the Stereo Vision Research Page of Middlebury College (http://www.middlebury.edu/stereo/) for comparison with the state-of-the-art results. Furthermore, Geometrix may incorporate parts of our research in their 3-D reconstruction products.

In terms of fundamental research, our future work will be along two major axes. First, on the augmentation of the tensor voting framework which will make it more efficient and capable of conveying more information. Second, on the proper incorporation of constraints specific to the stereo problem into the framework. In the short term, we will focus on the integration of curvature information into the framework and on the development of more efficient and accurate structure extraction techniques using dense Tensor Voting and a multi-resolution scheme. We will also work to overcome our current need for calibration information and on the development of a more time-efficient Tensor Voting engine. Finally, we aim at a Tensor Voting framework that will operate in multiple scales and will be able to handle extremely difficult cases of perceptual grouping some of which arise from complicated stereo sequences. We anticipate that the latter goal will require the most time and effort, but will stand as a significant milestone of our research.

9. Member Company Benefits

This approach, once successfully implemented, represents a significant improvement over the existing state-of-the-art. It can be incorporated directly into one of the company's product lines, which involves the creation of 3-D models from images. The focus of our future research will be on solving computer vision problems for which no satisfactory solutions exist in the literature.

10. References

[1] M.S. Lee and G. Medioni, "Inferring Segmented Surface Description from Stereo Data", Proc. CVPR, pp. 346-352, 1998.
[2] M.S. Lee and G. Medioni, "Grouping ., -, ->, O-, into Regions, Curves, and Junctions", CVIU, vol. 76, no. 1, pp. 54-69, 1999.
[3] M.S. Lee, G. Medioni, and P. Mordohai, "Inference of Segmented Overlapping Surfaces from Binocular Stereo", IEEE Trans. on PAMI, vol. 24, no. 6, pp. 824-837, June 2002.
[4] G. Medioni, M.S. Lee, and C.K Tang, "A Computational Framework for Feature Extraction and Segmentation", Elsevier, 2000.
[5] P. Mordohai, G. Medioni, and M.S. Lee, "Inference of Segmented Overlapping Surfaces from Binocular and Multiple-View Stereo", 3rd Workshop on Perceptual Organization in Computer Vision, 2001.
[6] P. Mordohai and G. Medioni, "Perceptual Grouping for Multiple View Stereo using Tensor Voting", International Conference on Pattern Recognition, vol. 3, pp. 639-644, Quebec City, Canada, August 2002.

[7] G. Medioni, P. Mordohai, and M. Nicolescu, "The Tensor Voting Framework", to appear in the Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics, Eduardo Bayro-Corrochano (editor), Springer-Verlag, 2003.
[8] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV, vol. 47, no. 1-3, pp. 7–42, April 2002.
[9] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer, "A Survey of Methods for Volumetric Scene Reconstruction from Photographs", Proc. International Workshop on Volume Graphics, 2001.