

# Real Time Stereo Based Obstacle Detection for UAV Threat Avoidance

**Jeffrey Byrne, Martin Cosgrove, Raman Mehra**

Scientific Systems Company, Inc.

500 West Cummings Park, Suite 3000, Woburn, MA 01801

{jbyrne,martin,rkm}@ssci.com

**Abstract:** *We present a system for UAV obstacle detection on embedded hardware based on Sarnoff Corp's Acadia I vision processor for 23Hz 640x480 binocular stereo and 10Hz minicut based recursive bipartition of an affinity graph. We briefly describe the system architecture, followed by performance results on simulated imagery, indoor and outdoor imagery, and flight experiments.*

**Keywords:** obstacle detection; collision avoidance; unmanned air vehicle; robotics; stereo; computer vision;

## Introduction

Unmanned Air Vehicles (UAVs) are envisioned as an integral part of future military forces. Large scale UAVs will perform autonomous tasks such as high-altitude reconnaissance, Close Air Support, Suppression of Enemy Air Defenses, and aerial refueling. Small scale UAVs will enable on demand Intelligence, Surveillance and Reconnaissance tasks including: “over the hill” reconnaissance, “perch and stare” surveillance, biological and chemical agent detection, precision strike missions, and battle damage assessment. These capabilities will provide unprecedented levels of support and intelligence to the military without endangering soldiers.

These tasks require that a UAV exhibit autonomous operation including *collision avoidance*. UAVs flying “nap of the earth” risk collision with obstacles whose position cannot be guaranteed as known a priori. UAVs must include situational awareness based on sensing and perception of the immediate environment to locate collision dangers and plan an appropriate avoidance path. Passive sensors for collision detection based on visual electro-optical (EO) or forward looking infrared (FLIR) promise a low power, covert solution that can enable micro scale UAV platforms and also provide an additional obstacle detection sensor for larger UAVs to complement active sensors for improved robustness [1].

Visual obstacle detection is difficult due to the inherent ambiguity of reconstructing three-dimensional (3D) scene structure from two-dimensional (2D) projected images. Reconstruction requires a solution to the *correspondence problem*, which establishes a match between features in two or more known views to enable triangulation for 3D scene reconstruction. Regions of low contrast, foreshortening distortion, specular reflections, periodic features and occlusions make the correspondence problem ambiguous and difficult [7]. Research in 3D reconstruction is widespread in the computer vision and robotics literature, so here we limit the scope to real time methods (>1Hz) appropriate for UAV obstacle detection.

Such approaches include biomimetic insect vision techniques using differential invariants of an optical flow field [2], structure from motion techniques [3], cluster tracking [4] and 2D/3D stereo evidence grids [5]. These approaches share the common difficulties of *sparseness* due to computational limitations causing missed detections and *local correspondence errors* causing false alarms.

This paper presents the Visual Threat Awareness (VISTA) system for visual obstacle detection. This system combines binocular stereo, image compression, image segmentation and region tracking for real time collision obstacle detection. The critical step in this system is real time, optimal image segmentation. We believe this can improve upon existing methods to reduce missed detections by focusing computation on regions likely to contain collision dangers, and reduce false alarms through region integration. We describe the system architecture, show qualitative segmentation performance on simulated, indoor and outdoor imagery, and demonstrate proof of concept in a flight experiment.

## VISTA System Architecture

Figure 1 shows a block diagram of the VISTA system. Imagery is captured from a calibrated, forward looking stereo camera pair and the stereo disparity map is computed using the Acadia I vision processor [6]. The imagery and disparity maps are compressed using the log-polar mapping to focus computation on regions likely to contain collision dangers and fused into an affinity graph representation using perceptual organization techniques. The affinity graph is recursively bipartitioned using a minimum s-t graph cut for an estimate of regions in the current field of view. Statistics are computed for each region, which provide a means for rejecting local correspondence errors. Those regions determined to be collision dangers are tracked and provide dynamic constraints for avoidance.

*Stereo:* Computational stereo is the process of extracting 3D scene structure from two or more images taken from distinct viewpoints [7]. This system uses the Acadia I vision processor for calibration, rectification, correspondence and reconstruction of 640x480 grayscale binocular imagery with a baseline of 0.5m at 23Hz [6]. The Acadia I solves for 32 disparity correspondence along epipolar scanlines using a 4-pyramid sum of absolute differences (SAD) approach, followed by left/right consistency check and SAD threshold. These checks discard those regions with poor correspondence due to low contrast, foreshortening distortion, specular reflections, periodic features and occlusions. The result is a *disparity*

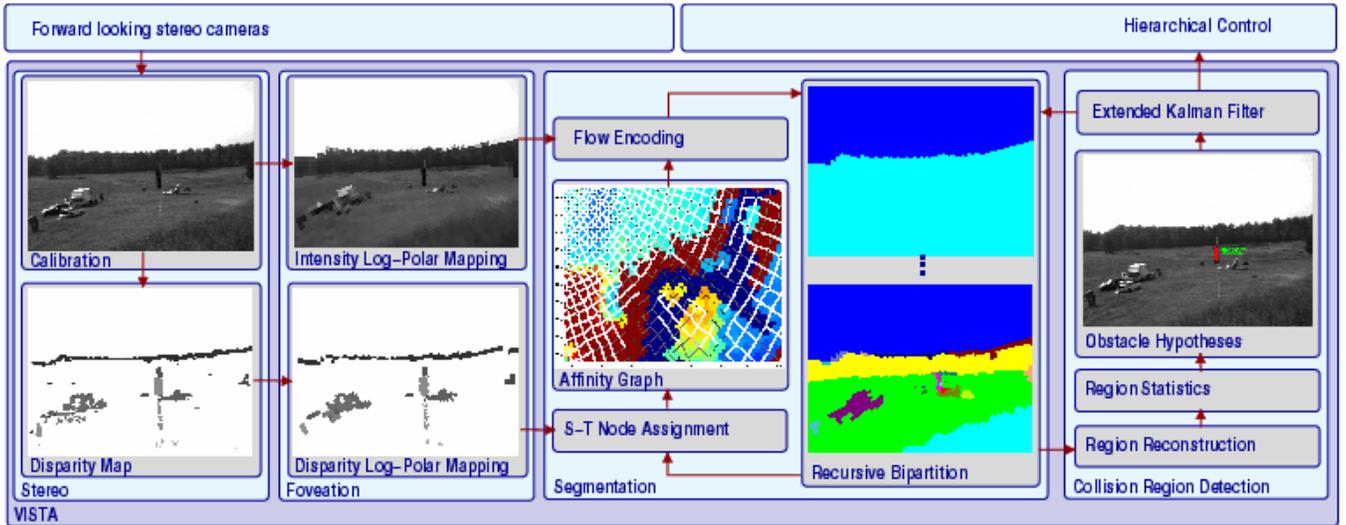


Figure 1. System block diagram and example imagery for a “pole” collision obstacle detection.

*map* used in 3D reconstruction for stereo triangulation. An example disparity map is shown in Figure 1, where dark encodes far, bright is close, and white is undefined.

*Foveation*: Log-polar mapping is an example of space variant computer vision [8] which refers to the smooth variation of resolution across a workspace such that the resolution is proportional to the log of the distance from the image center. This image representation provides the benefit of *foveation*, where imagery exhibits a high resolution central region or *fovea* whose resolution falls off with the log of the distance from the fovea center. In the context of obstacle detection, foveation provides an appropriate tradeoff between resolution and compression for real time performance, effectively focusing available computational resources on those spatial image regions that are likely to contain collision dangers. Example foveation using the log-polar mapping is shown in Figure 1.

*Segmentation*: Image segmentation is the problem of labeling regions in an image to form hypotheses of visual structure. One approach to segmentation is based on *perceptual organization*, which uses human perception inspired Gestalt criteria, such as proximity, common fate, closure, similarity, symmetry and continuity to group features. This approach permits the extraction of visual structure or *object hypotheses* with minimal domain knowledge, minimal assumptions and therefore minimal restrictions [9].

A graph theoretic approach to perceptual organization represents an image as an *affinity graph*. A graph  $G=(V,W)$  is defined by an ordered set of vertices or nodes  $V$  with  $|V|=N$  elements.  $W$  is an  $N \times N$  symmetric edge weighting matrix whose elements  $W_{ij}$  correspond to a measured relationship between the  $i^{\text{th}}$  and  $j^{\text{th}}$  nodes. In this application, images are abstracted to nodes using the log-polar mapping, and the  $M$  edge weights encode *affinity* between nodes with radial and angular connectivity in the log-polar mapping. Affinity is measured using the

perceptual organization heuristic of similarity as modeled by *feature smoothness*:

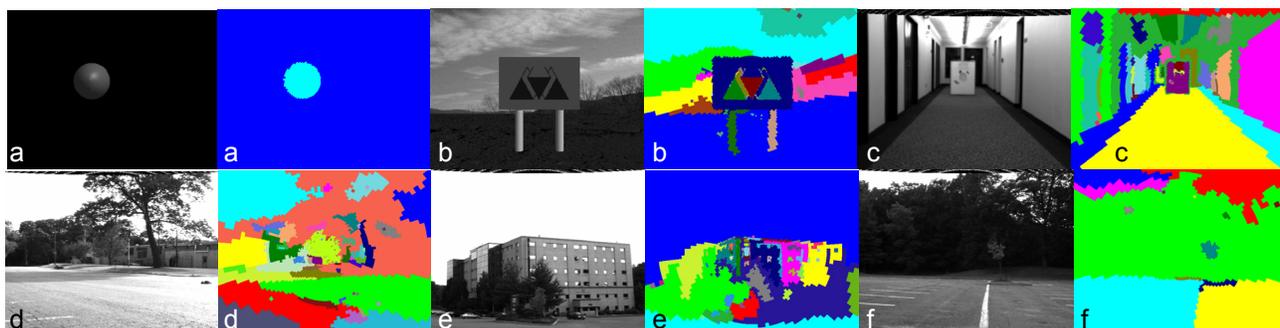
$$W_{ij} = \sum_{d=1}^D \alpha_d \left( 2 \exp \left( \frac{-\max((m_d^i - m_d^j), \mu_d)^2}{2\sigma_d^2} \right) - 1 \right) \quad (1)$$

for feature measurements  $m_d^i$ ,  $m_d^j$ , parameters for the  $d^{\text{th}}$  feature smoothness  $\mu_d$ ,  $\sigma_d$ , and feature weight  $\alpha_d$ . For this application,  $D=2$ ,  $\alpha_d=0.5$ ,  $\mu_d=2$ ,  $\sigma_d=30$  and features include intensity, disparity and connectivity. This affinity model is related to the approach in [10]. A visualization of a part of an affinity graph overlaid on a false color section of a foveated intensity image is shown in Figure 1.

Image segmentation requires a partitioning of the affinity graph to group features. A  $k$ -partition of the affinity graph  $G$  is a separation of the nodes  $V$  into  $k$  disjoint classes such that a cut measure  $N$  is minimized [11]. In this application, a  $k$ -partition is computed recursively using a minimum  $s$ - $t$  cut of a network flow graph. Edge weights  $W_{ij}$  are interpreted as flow capacities and certain distinguished nodes  $s$  and  $t$  are interpreted as terminal nodes. The minimum  $s$ - $t$  cut defines a membership matrix  $X$  which separates  $V$  into two disjoint sets  $V_s$  and  $V_t$  such  $s \in V_s$  and  $t \in V_t$  and the cut

$$N(X) = \sum_{i \in V_s, j \in V_t} W_{ij} \quad (2)$$

is a global minimum. Using the Ford-Fulkerson theorem, it can be shown that a solution that minimizes equation 2 is equivalent to a solution of a maximum network flow (maxflow) problem [12], for which there exist efficient polynomial time solutions. A  $k$ -partition is found by recursively bipartitioning with stopping criterion  $V_s = \emptyset$ . In this application, we use a new augmenting flows maxflow algorithm by Boykov and Kolmogorov [12] that exhibits polynomial computational complexity  $O(MN^2)$ . Terminal nodes  $s$  and  $t$  are assigned from disparity



**Figure 2.** Qualitative segmentation results for synthetic, indoor and outdoor ground based imagery. (a) synthetic sphere, (b) synthetic sign (c) indoor hallway, (d) large tree, (e) building and (f) small tree.

smoothness violation in equation 1, and from prior segmentation results. Flow encoding is introduced to avoid single node groups common with the mincut [13]. Example segmentation results are shown in Figure 2, where each of  $k$  partitions in the solution is displayed as a constant, randomly chosen color.

*Collision Region Detection:* Each region from image segmentation is 3D reconstructed using standard stereo triangulation to the mean region disparity. Statistics for each region are computed including disparity variance, and valid to invalid disparity ratio, and those regions without adequate statistics are discarded. The remaining regions that fall within a cylindrical collision volume are defined as obstacle hypotheses. Obstacle hypotheses states are represented by a bounding ellipse, which are then tracked using the standard formulation of an extended Kalman filter (EKF) with known linearized platform motion.

### Flight Experiment Results

Figure 3 shows results from the first four flight experiments of the VISTA system. Due to the risk involved in flying an aircraft on an obstacle collision path, these first experiments focused on quantifying position estimation for single obstacle scenarios only.

The flight experiments were performed on the Georgia Tech GT-Max autonomous helicopter platform [14], outfitted with the VISTA flight computer and stereo cameras. Flights 1-2 had the helicopter autonomously approach a “sign” obstacle, which was a 40”x30” piece of white foamcore mounted on the top of a 21’ tall, 0.75” diameter pole. Flights 3-4 replaced the “sign” with a “pole” obstacle, which was a 90”x20” piece of black foamcore representing the top section of a 20” diameter telephone pole. The helicopter approached the obstacles at a constant velocity and altitude, with variable heading (north/south or east/west), forward speed and ambient lighting for each flight.

Figure 3 shows example imagery from flights one and four, overlaid with the obstacle detection results shown as a green ellipse. Notice that the imagery from flight four also correctly detected a second collision obstacle in the background. This obstacle was an occluded gantry pole and wire at a distance of 140ft.

The plots in Figure 3 show tracking error of the obstacle centroid given the known ground truth obstacle location, overlaid with the predicted upper bound error from stereo range resolution. The predicted upper bound error reflects the nonlinear range resolution of stereo due to pixel quantization. The uncertainty in range  $\Delta r$  for a single disparity is proportional to the square of the range  $r$  given known baseline  $B$  and focal length  $f$  in pixel units:

$$\Delta r = \frac{r^2}{Bf} \quad (3)$$

The predicted error in equation 3 is an upper bound on the obstacle detection accuracy given known helicopter position from the onboard inertial navigation solution. These plots show that the tracking error remains close to this theoretical uncertainty, while occasionally achieving better performance due to tracking filter averaging. The table in Figure 3 shows the root-mean-squared (RMS) tracking error of the obstacle in a local inertial frame coordinate system over the entire flight, and the RMS deviation of the tracking error from the predicted upper bound. Runtime performance for each flight ranged from 5Hz-10Hz, due to scene complexity affecting the total number of regions  $k$  of the recursive bipartition.

### Conclusions

This paper demonstrated proof of concept in a flight experiment for the VISTA obstacle detection algorithm. Future experiments will explore autonomous avoidance, and multi-obstacle scenarios to quantify false alarm and missed detection rate in a complex environment.

### References

1. Bhanu, B., B. Roberts, D. Duncan, and S. Das. A system for obstacle detection during rotorcraft low altitude flight. *IEEE Transactions on Aerospace and Electronic Systems*, 32(3):785-897, July 1996.
2. Ancona, N., T. Poggio. Optical Flow from 1-D Correlation: Application to a simple Time-To-Crash Detector. *Int. J. Computer Vision*, (14)2, 1995
3. Tomasi, C. and T. Kanade, Factoring Image Sequences into Shape and Motion, *IEEE Workshop on Visual Motion*, p21-28, 1991

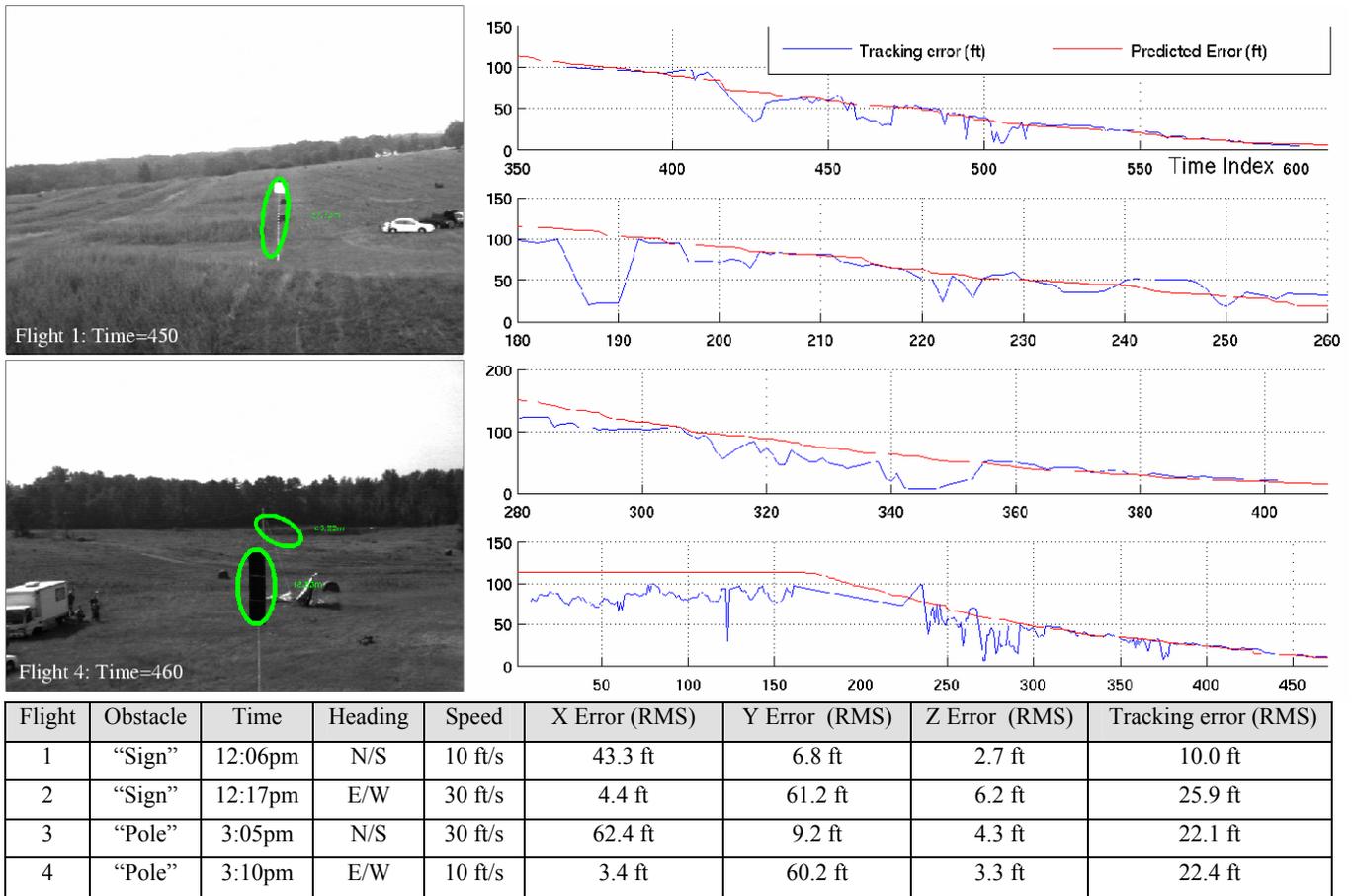


Figure 3. Flight experiment obstacle detection performance. Flight 1 (top) - Flight 4 (bottom)

4. Gandhi, T., M-T. Yang, R. Kasturi, O. Camps, L. Coraor, J. McCandless, Detection of Obstacles in the Flight Path of an Aircraft. *IEEE Transactions on Aerospace Electronics* 39(1):176-191, 2003
5. Moravec, H., Robot Spatial Perception by Stereoscopic Vision and 3D Evidence Grids, CMU-RI-TR-96-34, www.seegrid.com, September 1996
6. Van der Wal, G., M. Hansen, and M. Piacentino, .The Acadia Vision Processor., *IEEE International Workshop on Computer Architecture for Machine Perception*, Padua, Italy, September 2000
7. Brown, M.,Z. Burschka, G.D.Hager, Advances in Computational Stereo, *IEEE Trans.Pattern Analysis and Machine Intelligence*,25(8),p.993-1008, 2003
8. Wallace, R., P. Ong, B. Bederson, and E. Schwartz, Space variant image processing, *International Journal of Computer Vision*, 13(1):71.90, 1994.
9. Sarkar, S., and K. L. Boyer. Perceptual organization in computer vision: A review and proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):382.399, 1993.
10. Shi, J. and J.Malik. Normalized cuts and image segmentation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 22(8):888-905, August 2000.
11. Byrne, J., A. Gandhe, R. Prasanth, B. Ravichandran, M. Huff, R. Mehra, A k-partition, graph theoretic approach to perceptual organization, *IEEE conference on integration of knowledge intensive multi agent systems (KIMAS)*, p.336-342, October 2003.
12. Boykov, Y., and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in computer vision. In *International Workshop on Energy Minimization Methods in Computer Vision*, (2134), p359-374, 2001.
13. Soundararajan P., and S.Sarkar, Investigations of measures for grouping by graph partitioning,. in *International Conference on Computer Vision and Pattern Recognition*, December 2001.
14. Johnson, E., and D. Schrage, .The Georgia Tech Unmanned Aerial Research Vehicle: GT-Max. in *AIAA Guidance, Navigation, and Control Conf.*, 2003.