SUAVE: Integrating UAV Video Using a 3D Model

Shafiq Abedin, Michael Lewis
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260 U.S.A.
sha33@pitt.edu, ml@sis.pitt.edu

Nathan Brooks, Sean Owens, Paul Scerri,
Katia Sycara
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15260 U.S.A.
nbb@andrew.cmu.edu, owens@cs.cmu.edu,
pscerri@cs.cmu.edu, katia@cs.cmu.edu

Controlling an unmanned aerial vehicle (UAV) requires the operator to perform continuous surveillance and path planning. The operator's situation awareness degrades as an increasing number of surveillance videos must be viewed and integrated. The Picture-in-Picture display (PiP) provides a solution for integrating multiple UAV camera video by allowing the operator to view the video feed in the context of surrounding terrain. The experimental SUAVE (Simple Unmanned Aerial Vehicle Environment) display extends PiP methods by sampling imagery from the video stream to texture a 3D map of the terrain. The operator can then inspect this imagery using world in miniature (WIM) or fly-through methods. We investigate the properties and advantages of SUAVE in the context of a search mission with 3 UAVs.

INTRODUCTION

Controlling an unmanned aerial vehicle (UAV) requires the operator to perform continuous surveillance and path planning, often making the operator's task tedious and mundane (Quigleyet al., 2004). For a solution scaling to multiple UAVs, this kind of one-to-one surveillance is not feasible. Also, the operator's situation awareness of the context degrades by multiplying the number of surveillance videos that must be viewed and integrated (Calhoun et al., 2005, Tsoet al., 2003).

The method of Picture-in-Picture display (PiP), a specialized solution for integrating UAV camera video (Draper et al., 2006, Hunn, 2005), has been proposed to solve the problem of integrating information in-context to maintain situation awareness (SA). In a PiP presentation, the operator's video feed is scaled and transformed so it may be viewed in the context of surrounding terrain eliminating the 'world-through-a-straw' effect (Woods, et al. 2002). The video feed is projected onto a map thus expanding the context of the operator and reducing the mental transformation and ambiguity of interpreting the video from a remote camera (Gugerty et al., 2001). These displays (Calhoun et al., 2005; Draper et al., 2006, Drury et al., 2006) typically provide a partial iconic view of the UAV revealing its position and orientation and a heading-up view of the map with video projection. As the UAV flies through the environment, the operator's view of the video moves with it, with surrounding areas of the map providing context. This type of tethered viewpoint has been shown to improve situation awareness and performance over ego-centric viewpoints in a variety of applications (Milgram et al., 1993; Nielsen et al., 2007; Wang et al., 2009) as well as PiP displays (Draper et al., 2006).

World in miniature (WIM) (Pausch et al., 1995) and fly-through (Bowman et al., 1997) model-inspection techniques offer an alternative approach for interacting with camera imagery in the context of a map. In WIM, also called world-in-hand, interaction the user can zoom, pan, or tilt a 3D model to inspect it. Allowing the user to fly-through an anchored model is the natural complement of WIM. With these methods developed for interaction with virtual environments and games the operator is allowed to concentrate on exploring and understanding the environment rather than focusing on the imagery and context of particular platforms, an orientation Alberts, et al. (1999) refer to as *network centric*. The operator's task becomes a simple visual search of a map without all the mental transformations and demands on memory needed to integrate current and past imagery from multiple UAVs. Currently, use of these techniques for UAV imagery (Kumar et al., 2001; Page, 1999) has been limited to access and exploitation of archival data.

Simple UAV Environment (SUAVE) is an experimental system being developed to investigate the use of model-inspection techniques to exploit real-time video feeds. One of the benefits of model-inspection based display is that temporal and spatial resolution can be traded off. If data is collected at high spatial resolution, then large regions can be searched and inspected closely but some data may be obsolete. If large areas must be surveilled for rapidly unfolding events, spatial resolution can be sacrificed and temporal resolution maintained by having the platforms cover larger areas at a higher frequency. This approach has favorable scaling effects for human-UAV interaction because adding UAVs acts either to improve the frequency at which imagery is updated (temporal resolution) or the spatial resolution at which it is collected without imposing extra load on the operator.

SUAVE and other model-inspection approaches are inherently asynchronous. While PiP displays provide a context for viewing a real-time video feed, SUAVE samples imagery from the video stream to apply textures to its map. Because simultaneity is lost, the user can no longer be guaranteed to see new events on screen as they occur. Viewing UAV video feeds directly or through PiP poses the same problem of unseen events but avoids confusion between new and old imagery. Where dynamic events are not the focus, as in searching for immobilized victims or other foraging tasks, asynchronous display types such as SUAVE are ideal. Figure 1 shows the interface and its elements.

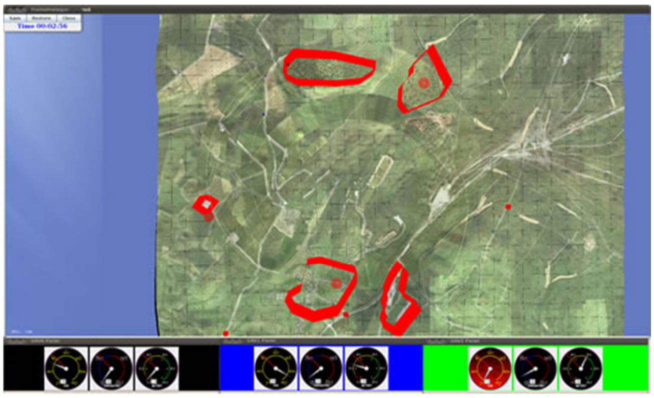

Figure.1 SUAVE interface with critical regions in red

The focus of our current research is on developing techniques that allow model-inspection displays, such as SUAVE, to be used effectively in dynamic environments.

Some predictable advantages of this approach are:

- An increase in temporal and spatial resolution with multiple UAVs without increasing task difficulty.
- A centralized mechanism that allows user to perform secondary tasks (i.e. path planning), potentially taking user preferences for priority and update rates into account.
- Efficient utilization of the data transmission rate by only requesting imagery from the highest priority UAVs or areas that has not been traversed or only traversing through areas of interest.
- Added model-inspection could increase situation awareness more than displays requiring the operator to follow a video stream while engaged in secondary tasks (Blinn et al., 1988, Cummings et al., 2005, Drury et al., 2006, Richer et al., 2006).
- When engaged in secondary tasks or, distracted, the operator can still recover missed targets because the updated imagery remains present in the 3d terrain model. The operators can inspect the terrain at their own leisure.

In this paper we test the hypothesis that an operator's situation awareness can be enhanced by an asynchronous 3d terrain model (SUAVE) in a dynamic environment.

SUAVE

In our version of WIM we create a 3D model with initial texture, get live video feed from the UAVs along with telemetry data, select the individual frames and paint them onto the terrain. Figure 2 shows the entire process.

In SUAVE we begin with a 3D terrain from satellite imagery or other previously acquired aerial data to provide geographical features. As the UAVs capture live video streams, individual frames are selected and projected onto the terrain replacing the old texture. Once the terrain is created georeferencing is used to map the triangles in the mesh. Then a list of texel points corresponding to these triangles is used to map the imagery onto the map. Along with the texture coordinates each texel point has 3D world space coordinates. For each video frame, visible texels are computed from the viewpoint of the UAV and then all the triangles that are outside the UAVs view frustum are culled. Then the triangles are projected onto the 2D plane and intersection test is used to reduce each triangle visible only by the UAV. Finally for each texel point, color is sampled from the projected location and then onto the texture.

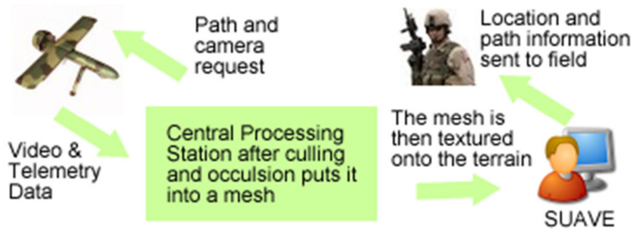

Figure.2 Illustration of the SUAVE system

The operator has the ability move freely in the miniature world with six degrees of freedom (6DoF).This gives the platform the versatility of:

- Giving user the ability to interact with the 3d model.
- Allowing them to inspect the world at their own pace
- Allowing the operator to prioritize their tasks rather than limiting them to fixed video frames and having to look at them in order to regain context.

We compare SUAVE to a video surveillance mode in which the user is required to synchronously monitor the video feed for all three UAVs. In contrast to this synchronous viewing model, the asynchronous 3D terrain model may relieve the operator of this load by breaking it into the aforementioned tasks. Prior work (Wang et. al. 2011) showed a comparison between synchronous and asynchronous displays for static targets in which asynchronous display have advantages in terms of the operator's accuracy in marking targets in the environment. This effect may carry over to dynamic environments of the type we are considering.

METHODS

VBS2

As a simulation platform we use VBS2, a game-based training platform for high fidelity virtual environments with the ability to change scenarios and operate vehicles (aerial vehicle in this case). This battlefield simulation has been used to run the UAVs for this experiment. The video feed and the telemetry data has been collected from this simulation and fed into SUAVE. We also set predefined paths for the targets and the UAVs with VBS2, as explained below.

Experimental Conditions

We designed two conditions for the experiment. One is the synchronous display of information in video feeds (Figure 3) and the other is the asynchronous 3D reconstruction based on SUAVE. We created a new scenario in VBS2 and flew three UAVs in the virtual battle space. Both conditions received information from identical video streams from the three UAVs. For the first condition (video feeds) we also took the telemetry data from VBS2 and created a mini-map. The operator can click on the mini-map to localize and mark targets.

Figure 3. Interface of the first condition (video feeds).

For the second condition we took all three video streams and telemetry data and fed it to SUAVE for rendering on the 3D terrain. The operator can click anywhere on the WIM and mark the targets. For both conditions we added a dial panel with three dials for each UAV for additional tasks. The dials simulated real life data and errors (i.e. turn red when low on fuel).

Participants and Procedure

12 participants were recruited form the University of Pittsburgh with no prior experience in robot control, although most of them are frequent computer users.

Participants read the description of both conditions and were instructed on how to control the camera view for the second condition followed by a 30 minute training session. The participant then spent 15 minutes for each condition. In the first condition the participants spent their time observing the three synchronous video streams along with a mini-map for context. For the second condition, the participants inspected a high-resolution image projected on the terrain map using video game-like fly-through control to move about the map.

In both conditions participants were instructed to mark a predefined target whenever they encountered it. We used situation awareness assessment techniques (Endsley, 1995) to evaluate the situation awareness at random intervals. These questions were concerned with the participant's general knowledge of the environment. At the end of each session, participants were asked to complete the NASA- TLX workload survey (Hart et al., 1988).

RESULTS

Data were analyzed using a repeated measures ANOVA comparing video stream with the SUAVE condition. Overall, in both conditions participants were successful in searching through the environment. Every mark a participant made for a target was compared to ground truth to determine whether there was in fact a target at the location. When targets were counted as successfully marked when within a 50 meters range, the result of ANOVA showed significant advantage for the SUAV condition ($F_{1,11}$= 19.186, p = .001). When considering a range of 100 meter, on average participants in the video stream condition successfully marked 2.75target while those in the SUAVE condition marked2.83 (Figure 4) without a significant difference between conditions ($F_{1,11}$= .009, p = .927).

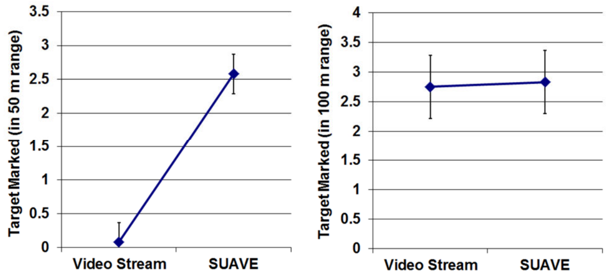

Figure 4. Targets marked correctly within a range of 50m (left) and 100m (right)

A mark made further than 100 meters away from any target or multiple marks for one target were always counted as *false positives*. Targets that were missed, but present in the video feed, and not marked were counted as *false negatives*. The number of false positive shows significant advantage for SUAVE condition than the video stream condition ($F_{1,11}$= 57.750, p < .001). However, the ANOVA result of false negatives showed no significant difference between the two conditions, $F_{1,11}$= .010, p = .923 (Fig. 5).

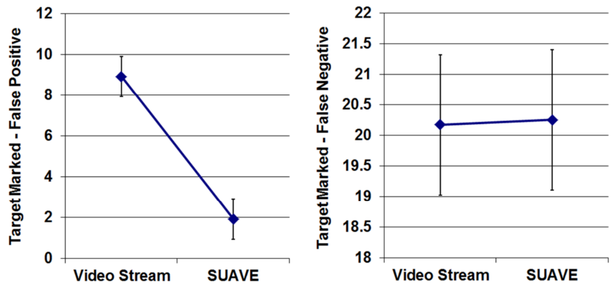

Figure 5. Marking errors of targets

The repeated measures ANOVA of the SA measure found a significant advantage in correct answers for participants in the SUAVE condition($F_{1,11}$= 10.000, p = .009).

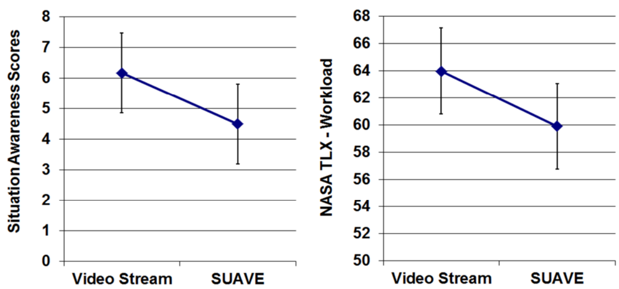

Figure 6. Situation awareness and workload

The full scale NASA-TLX workload measure, however, revealed no workload advantage ($F_{1,11}$= 1.074, p = .322 (Figure 6).

DISCUSSION AND CONCLUSION

Our experiments revealed that users were able to identify targets more accurately in the asynchronous condition (SUAVE). This is in contrast to the synchronous (video feed) condition where information is presented to the operator as it is acquired. Our experiment also revealed that in the streaming condition we had higher number of false positives since the operator had less time and opportunity to inspect the terrain and identify the target whereas in the SUAVE condition there were fewer and more accurate markers. The situation awareness and workload measures yielded no significant results but results for false positives indicate that the asynchronous condition may, in fact, present the relevant information with greater spatial resolution and better context for supporting situation awareness.

The current system presents some challenges that can be improved. For example, some users reported that their experience could be improved with better interface controls for the fly-through. As our experience with such systems improves, these initial shortcomings will be overcome and we hope to be able to fully exploit the advantages in performance and scalability that our experiment suggests may be possible.

ACKNOWLEDGMENT

This research has been sponsored in part by AFOSR FA955008-10356 and ONR Grant N0001409-10680

REFERENCES

Blinn, J.(1998). Where am I? What am I looking at. IEEE Computer Graphics and Applications, 8(4):76-81, 1988

Bowman, D., Koller, D.; Hodges, L. (1997). Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques, Virtual Reality Annual International Symposium, 1997., IEEE 1997 , 45-52.

Calhoun, G., Draper, M. and Nelson, J. (2006). Advanced Display Concepts for Uav Sensor Operations: Landmark Cues And Picture-In-Picture. In Human Factors and Ergonomics Society Annual Meeting Proceedings, volume 50, (pp.121–125). Human Factors and Ergonomics Society.

Calhoun, G., Draper, M., Abernathy, F.,Patzek, M. and Delgado, F.(2005).Synthetic vision system for improving unmanned aerial vehicle operator situation awareness. Enhanced and Synthetic Vision, 5802(1).

Cummings, M. & Mitchell, P.(2005). Management of multiple dynamic human supervisory control tasks for UAVs.In Human Computer Interaction International Human Systems Integration Conference, 2005.

Draper, M.H., Calhoun, G.L., Nelson, J.T., & Ruff, H.A. (2006). Evaluation of synthetic vision overlay concepts for UAV sensor operations: landmark cues and picture-in-picture.Proceedings of the Human Factors and Ergonomic Society 50th Annual Meeting, San Francisco, CA.

Drury, J. L., Richer, J., Racklie, N. & Goodrich, M. A. (2006). Comparing situation awareness for two unmanned aerial vehicle human interface approaches. In Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics,.

Drury, J., Riek, L. &Racklie, N.(2006). A decomposition of UAV-related situation awareness.In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, (pp. 88-94). ACM New York, NY, USA.

Endsley, M. (1995). Measurement of situation awareness in dynamic systems, Human Factors, 37(1), 65-84.

Gugerty, L., & Brooks, J. (2001). Seeing Where You Are Heading: Integrating Environmental and Egocentric Reference Frames in Cardinal Direction Judgements. Journal of Experimental Psychology: Applied, 7 (3), 251-66.

Hart, S., and Staveland, L. 1988). Development of a multi- dimensional workload rating scale: Results of empirical andtheoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload, 139-183. Amsterdam, The Netherlands: Elsevier.

Hunn, B. (2005). The Human Challenges of Command and Control with Multiple Unmanned Aerial Vehicles.Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.

Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hanna, A. Pope, R. Wildes, D. Hirvonen, M. Hansen, and P. Burt. (2001). Aerial video surveillance and exploitation.Proceedings of the IEEE, 89(10).

Milgram, P., Zhai, S., Drascic, D., Grodski, J. (1993).Applications of augmented reality for human-robot communication, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 1467-1472.

Nielsen, C., Goodrich, M., & Ricks, B. (2007). Ecological Interfaces for Improving Mobile Robot Teleoperation. IEEE Transactions on Robotics and Automation. Vol.23, No 5, 927-941.

Page, T. (1999). Incorporating scene mosaics as visual indexes into uav video imagery databases. Master’s thesis, Air Force Institute of Technology, Wright-Patterson AFB.

Partala, T. (1999). Controlling a Single 3D Object: Viewpoint Metaphors, Speed and Subjective Satisfaction. Human-Computer Interaction. In Proceedings of INTERACT'99, Edinburgh, UK, IOS Press. 486-493.

Pausch, R., Burnette, T., Brockway, D., &Weiblen, M. (1995). Navigation and locomotion in virtual worlds via flight into hand-held miniatures, Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH’95), 399- 400.

Pitman, A., Humphrey,C., & Adams, J..A Picture-in-Picture Interface for a Multiple Robot System.

Quigley, M., Goodrich, M. & R. Beard, R. (2004). Semi-autonomous human-uav interfaces for fixed-wing mini-uavs. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. VOL. 3.

Tittle, J.; Roesler, A., & Woods, D. (2002).The remote perception problem, Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 260-264.

Tso, K., Tharp, G., Tai, A., Draper, M., Calhoun, G. and Ruff., H. (2003). A human factors testbed for command and control of unmanned air vehicles. In Digital Avionics Systems Conference. Vol. 2..

Wang, H., Kolling, A., Abedin, S., Lee, P., Chien, S., Lewis, M., Brooks, N., Owens, S., Scerri, P. &Sycara, K. (2011) Scalable target detection for large robot teams, Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction.

Wang, W. &Milgram, P. (2009). Viewpoint Animation With a Dynamic Tether for Supporting Navigation in a Virtual Environment. Human Factors 51(3): 393-403