Real Time Video Overlays

Donald Madden, Andrew Scanlon, Yunxian Zhou, Tae Eun Choe,* Martin Smith†
ObjectVideo Inc.

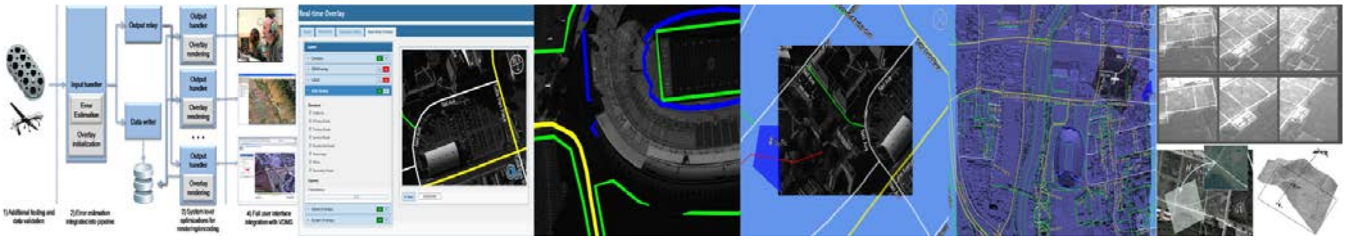

Figure 1: *Real Time Video Overlays (a) Overall architecture (b) Concurrent multi-user web interface (c) Occlusion handling (d) Expanded context (e) Virtual perspective and LiDAR projection (f) Sensor metadata error correction*

Abstract

We introduce a distributed augmented reality framework for aerial video which uses CPU/GPU acceleration to correct sensor metadata errors, create a geo-referenced scene model registered to the video, overlay important data, and stream to multiple web clients in order to improve situational awareness during real-time missions.

Keywords: Unmanned aerial vehicles, mobile augmented reality, situational awareness, sensor pose estimation, distributed systems

1 Introduction

Unmanned aerial vehicles (UAVs) are critical to provide intelligence, surveillance and reconnaissance capabilities. Recently, UAV operations and deployment have increased dramatically for military and commercial missions. A wealth of geo-referenced data is available to aid users, such as points of interests, roadways, airspace symbology, and LiDAR. Due to limited camera viewing angles and resolution, users lack sufficient context to make critical decisions. Additionally, the camera angles can make even familiar objects difficult to recognize and understand. Video overlays may suffer from sensor jitter, drift and incorrect initial states. Also, a multi-user concurrent web interface is necessary for remote monitoring. Therefore, we developed a distributed real-time video overlay framework to overlay arbitrary datasets onto video feeds for multiple users.

2 The Video Overlay System

The system provides a streaming server that processes the incoming UAV video feeds, corrects for metadata error, overlays the geographic data onto the video feed, and streams that video to multiple HTML5 web UIs. The server renders overlays dynamically for each stream client. To maximize system performance, feature matching, sensor pose optimization, video codecs, and rendering engine all utilize multi-core CPU/GPU accelerations. The system is scalable to increasing video data volumes, concurrent users, and future capabilities. The core components are: *1) VESPA Framework*, the hardware accelerated streaming data management platform that processes and disseminates video and metadata streams, manages connections, handles data storage and retrieval, and provides custom processing of media streams. The framework provides a flexible distributed workflow architecture where multiple low-level processing modules are coupled at runtime to form a single continuous processing chain. *2) Concurrent Multi-User Web Interface* for users to configure overlay layers or rendering preferences, customize layer appearance, and view the resulting video overlays. *3) Overlay Asset Manager* to provide centralized access to external datasets, reference data, and custom plug-in implementations. *4) Overlay SDK* to enable rapid integration of arbitrary datasets by providing third-party integrators with common rendering elements to develop overlay layers for any dataset. *5) Sensor Metadata Error Correction* module to correct for inaccuracies, drifts, and temporal shifts in the incoming sensor telemetry data. It integrates multiple components including frame input/output/sync, sensor gap filter, coordinate conversion and projection, coarse and robust registrations [Szeliski 2011], motion estimation, triangulation and bundle adjustment. *6) Overlay Engine*, responsible for handling incoming frames, metadata, and rendering results. It maintains all assets and handles the management of plug-ins. A virtual sensor coordinate system is established using the corrected metadata, to which 2D/3D geo-overlay assets are transformed and displayed. Proper occlusion modeling provides better understanding of video elements. LiDAR data is used to construct a depth map and determine the visibility of overlays versus the video.To mitigate the soda-straw effect, the viewing area is expanded and a virtual sensor can be defined at arbitrary locations and viewing angles to the video and map tiles.

3 Future Work

In the future, graphical optimizations for large area display will be developed. The asset manager will provide a centralized data service. Predictive performance modeling will be integrated to correct sensor metadata. Performance evaluation will be conducted with a plug-in based workflow processing framework.

References

SZELISKI, R. 2011. *Computer Vision: Algorithms and Applications*. Springer.

*e-mail:{dmadden, ascanlon, szhou, tchoe}@objectvideo.com
†e-mail:martinws@gmail.com

SIGGRAPH 2014, August 10 – 14, 2014, Vancouver, British Columbia, Canada.

ACM 978-1-4503-2958-3/14/08