

New system performs persistent wide-area aerial surveillance

Arjun Nagendran, Don Harper, and Mubarak Shah

Visual sensors and an inertial navigation system mounted on a helium balloon can collect high-definition video that is synchronized with metadata.

Computer-vision algorithms for automatic-target detection and tracking in aerial imagery are being developed to assist in surveillance. Airborne platforms, including autonomous unmanned vehicles, deployed in sensitive areas play an important role in activity-monitoring and threat-detection, and are increasingly being used for intelligence and reconnaissance missions. Often, a trade-off between the field-of-view and the imagery resolution is required to focus on a specific activity on the ground. This determines the ground-sampling distance (GSD), or the number of pixels on an image target. Since the performance of any automated detection or tracking algorithm is directly proportional to the GSD, several solutions to persistent wide-area surveillance systems at high resolution are being developed. For example, the Defense Advanced Research Project Agency's Autonomous Real-time Ground Ubiquitous Surveillance-Imaging System (ARGUS-IS)¹ and Geospatial Systems Inc.² are attempting to provide high-resolution imagery over a wide field-of-view.

The performance of existing computer-vision algorithms is typically evaluated on certain standard datasets that have low resolution and a narrow field-of-view. To develop algorithms for wide-area surveillance systems and high-resolution imagery acquisition, it is necessary to simulate and get representative datasets from aerial platforms. Additionally, it is particularly useful to obtain information about the sensor's location and orientation (referred to as metadata) which can be valuable in post-processing. The combined need for metadata and high-resolution imagery also requires sufficient disk-storage space, which adds to the aerial platform's payload and eliminates the possibility of using small-scale, fixed-wing aircraft or rotorcraft for data collection. At the University of Florida's Compute Vision Lab, we have developed a platform to overcome these shortcomings.



Figure 1. The 13' Kingfisher Aerostat helium balloon.



Figure 2. The payload platforms—non-gyroscope-stabilized (left) and gyroscope-stabilized (right).

Our aerial platform consists of a 13' Kingfisher Aerostat helium balloon (see Figure 1) that is designed to provide a stable

Continued on next page

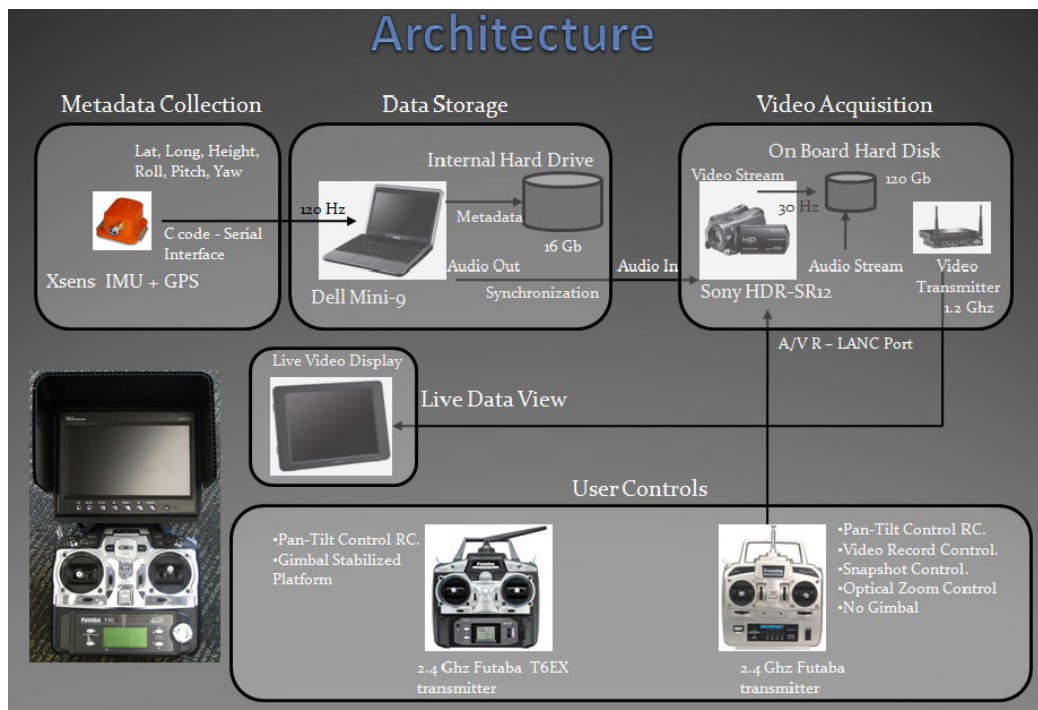


Figure 3. Communications architecture of a surveillance system. IMU=inertial measurement unit. GPS=global positioning system. RC=remote control. LANC=local application control bus system. A/V R=audio-video recording.

operating platform in winds up to 50mph. It can lift a rated payload of 34lbs while assuring an even weight distribution and line tension across the entire envelope. The balloon is tethered to a vehicle with an electric winch and can be controlled to reach the required altitude. The vehicle's movement helps simulate motion in the aerial videos at reasonable velocities depending on the line tension.

The payload platform attached to the balloon consists of multiple sensors. Two pan-tilt-roll units are typically deployed (see Figure 2), a non-gyroscope-stabilized platform capable of stable operation at low wind speeds and a Tazor 3X Mark IV gyroscope-stabilized gimbal platform stabilized by Futaba gy401 gyroscopes. One of these is suspended from the payload platform. It can incorporate one of several visual sensors on its base. The visual sensors include a Sony HDR-SR12 camera that weighs about 1.4lbs and can record 1920×1080i (interlaced field) high-definition video at 30fps (frames per second) and a Sanyo Xacti VPC FH1 camera that weighs approximately 0.8lbs and can record full HD video of 1920 × 1080p (frame progressive) at 60fps.

An integrated inertial measurement unit (IMU) and global positioning system (GPS) from Xsens Technologies BV are mounted on the pan-tilt-roll unit, with its local axes aligned with

those of the camera, but offset by a calculated amount. This unit provides the necessary metadata during collection, including the longitude, latitude, height, roll, pitch, and yaw of the visual sensor. A Dell Inspiron mini-9 netbook computer that weighs 2.28lbs serves as the communication link with the IMU and GPS, and is mounted on the payload platform.

Figure 3 shows the communications architecture of our system. The Xsens IMU and GPS are interfaced through a serial communication protocol to the netbook, whose hard disk stores the metadata. High-definition video is recorded directly to the camera's storage device in raw, uncompressed format. An on-board 1.2GHz video transmitter streams the recorded video to a live display on the ground. The display is integrated with one of two Futaba 2.4GHz remote control (RC) units, depending on the payload platform used. The RC units can control the camera's roll, pitch, and yaw while simultaneously viewing the scene the camera is visualizing. The HDR-SR12 camera is equipped with a local application control bus system port interfaced with the transmitter's joystick to control the camera's zoom level, switch-to-snapshot mode, or video recording, depending on the joystick's movement.

Continued on next page

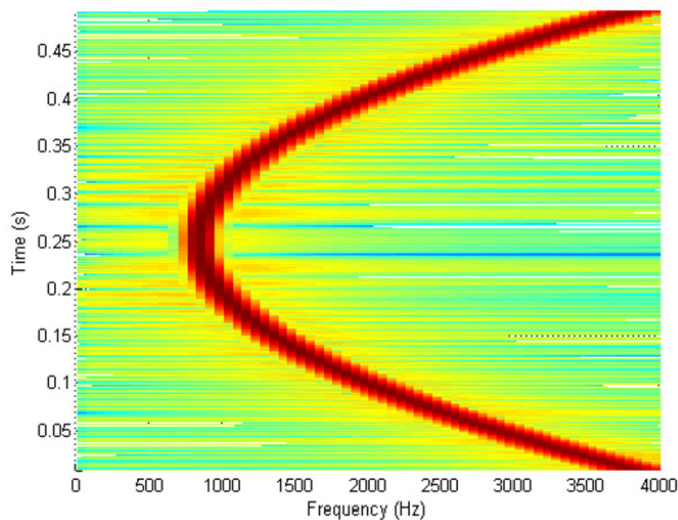


Figure 4. Spectrogram of a quadratic chirp signal.

Since the video and metadata are recorded independently, the two must be synchronized. We accomplish this using a quadratic chirp signal that has an easily recognizable spectrogram (see Figure 4). The netbook generates the chirp at regular intervals and sends them to the video camera's audio input port. The time at which this chirp is generated is recorded in relation to the metadata using the IMU's clock. The system analyzes the audio waveform from the recorded video, and the common time-stamp across the metadata and the generated chirp signal are automatically tallied by an algorithm. Each frame is synchronized with the metadata, which is then verified by generating a synthetic video³ from the sensor model using a reference image of the location where the video was taken.

The aerial platform is extremely useful in collecting long duration videos of 2–3 hours at a high resolution with metadata, and can hover over any region of interest, including stadiums, congested traffic junctions, and parking lots. The video obtained is being used to address several computer-vision challenges, including activity recognition, target tracking, detecting unusual behavior, and crowd analysis, all from the perspective of the wide-area persistent surveillance systems.^{1,2} We are now working on integrating the high-resolution aerial videos from this platform, roof-top cameras, and ground cameras and investigating potential multiple camera networks.

Author Information

Arjun Nagendran

Computer Vision Laboratory
University of Central Florida (UCF)
Orlando, FL

Arjun Nagendran, postdoctoral associate, holds a PhD in robotics from the University of Manchester, UK. He has been involved in the Ministry of Defense Grand Challenge (UK, 2008) and projects with the Florida Department of Transportation, and specializes in control and landing mechanisms for unmanned aerial vehicles.

Don Harper and Mubarak Shah

University of Central Florida (UCF)
Orlando, FL

Don Harper is the system administrator of the College of Electrical Engineering and Computer Science. He was the group leader for UCF's KnightRider team for the DARPA grand challenge that made it to finals. He has experience with the Bergen R/C Tazor 800 helicopter plus Tazor 3X three-axis, gyroscope-stabilized camera mount to capture aerial videos, and is responsible for building and flying UCF's aerial vehicles.

Mubarak Shah, founding director of the Computer Vision Lab at UCF, is a fellow of the Institute of Electrical and Electronics Engineers, SPIE, International Association for Pattern Recognition, and the American Association for the Advancement of Science. He co-authored three books, *Automated Multi-Camera Surveillance: Algorithms and Practice* (2008), *Video Registration* (2003), and *Motion-based Recognition* (1997), all by Springer. In 2006, he was named a Pegasus Professor, the highest award given to a UCF faculty member.

References

1. Brian Leininger, *A next-generation system enables persistent surveillance of wide areas*, SPIE Defense & Security. Accessed 24 June 2010., 2008. <http://spie.org/x23645.xml?ArticleID=x23645>, doi:10.1117/2.1200803.1112
2. Wide-aerial persistent surveillance for IED detection and homeland security using direct geo-referenced airborne imagery. <http://www.geospatialsystems.com/custom-solutions/persistent-surveillance/>. Accessed 24 June 2010.
3. Sample synthetic video generated at the Computer Vision Lab, University of Central Florida. <http://server.cs.ucf.edu/~vision/harris/results.avi>. Accessed 24 June 2010.