

Detection and tracking of objects from multiple airborne cameras

Mubarak Shah, Asaad Hakeem, and Arslan Basharat

Novel computer vision techniques using multiple commercial off-the-shelf (COTS) sensors can be used on mobile platforms to efficiently detect and track objects over large areas.

Video sensor networks play a vital role in unattended wide area surveillance. Most of the computer vision research in this area deals with networks of stationary electro-optical sensors. But recently there has been increased interest in networks using sensors on mobile platforms such as mobile robots, all-terrain vehicles, and unmanned aerial vehicles (UAVs). Here, we present novel computer-vision techniques for automatic object detection and tracking in mobile sensor networks. We use multiple commercial off-the-shelf (COTS) sensors that enable monitoring over large areas. The effectiveness of the object detection and tracking framework is demonstrated using aerial videos from multiple mobile sensors aboard UAVs (a sample UAV is shown in Figure 1).

The goal of an effective video surveillance system is to detect objects in an area of interest and to find their correspondence across many frames. There are several issues inherent to the problem, such as rapidly changing lighting conditions (e.g. due to cloud cover), shadow, occlusion, and entry/exit of objects. The small size of objects taken through aerial imagery is another issue, making it difficult to detect and track through varying appearance and frequent occlusions. To ease this problem, we use object appearance, shape, and motion models to detect and track objects from a single UAV. Once the objects are tracked successfully, we apply geometric similarity between object trajectories across UAVs to obtain consistent labeling (global object correspondence).

Conventional tracking methods that exploit appearance or position similarity are not suitable because of significant camera motion and object positional variation.

So, we have developed the COCOA system,¹ which uses three modules to achieve object detection and tracking in a single



Figure 1. Shown is an UAV used in our experiments.

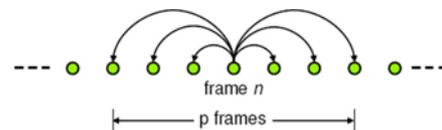


Figure 2. Frame difference analysis of each frame with respect to each of its p neighbors.

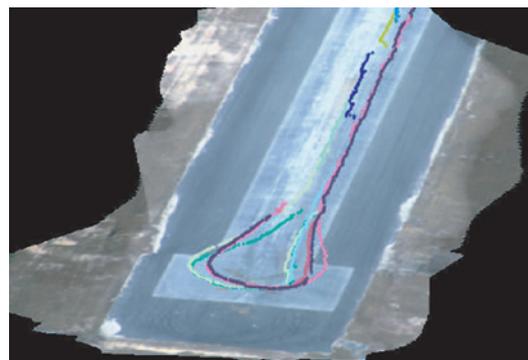


Figure 3. The results of the COCOA system are shown above. The generated tracks (shown in a different color for each moving object) are overlaid on the mosaic (landscape) and show global trajectories.

Continued on next page

UAV. The first module applies ego-motion compensation to all video frames so that the global motion of the UAV is compensated for. This allows moving objects to be detected in the area under surveillance. With this approach, a Harris corner detector is used to find feature points in the image. A small neighborhood around these is then matched by applying correlation to regions around feature points in other frames. The tentative matches are then filtered using RANSAC (the random sample consensus algorithm) to find the feature correspondence that has the best fit homography (geometric relationship). After the global motion is compensated for, the second module detects the motion of independently moving objects such as cars, trucks, tanks, or other objects using a frame-differencing scheme.

The object evidence is accumulated by computing the frame difference of each frame with respect to its p neighboring frames as shown in Figure 2. The log-evidence is accumulated for each frame and used to distinguish foreground pixels from background. Due to the motion of the objects, the boxes that show object position do not fit tightly.

The frame differencing provides a good initialization for object detection and tracking, the third module of the COCOA system. The initial object boundaries are refined using a level set-based segmentation approach that produces a contour that tightly detects independently moving objects. These are then tracked across frames until it exits the UAV field of view. To ensure the adaptivity of the meanshift tracker used, the object template is updated and maintained using the results of motion detection and level-set segmentation. For occlusion handling, we maintain shape and velocity as well as object appearance models.

We tested the object detection on 40 videos from a UAV, while the object detection and tracking was tested on 20 videos. Across all frames with detection rates were over 90% and false alarm rates were 1–3%. The result of one video is shown in Figure 3.

The COCOA system is run separately on each UAV video to generate respective object trajectories. Each in a given UAV is then matched to the trajectories in the others. This is achieved using a maximum likelihood estimation of the trajectory-similarity measurement, based on a cost function using algebraic and geometric distances between trajectories. Since trajectories of the same object acquired by two different UAVs are generated from a common 3D trajectory on the ground plane, we can compute a homography between these two trajectories using a direct linear transform algorithm. The likelihood of correspondence between two object is based on the re-projection distance (itself based on inter-trajectory homography) between them.

In the case of two moving UAVs, we can find the global correspondence using maximum matching of a complete bi-partite graph. In such a graph, all the nodes in a bi-partition represent the trajectories from a particular UAV. The edge-weights

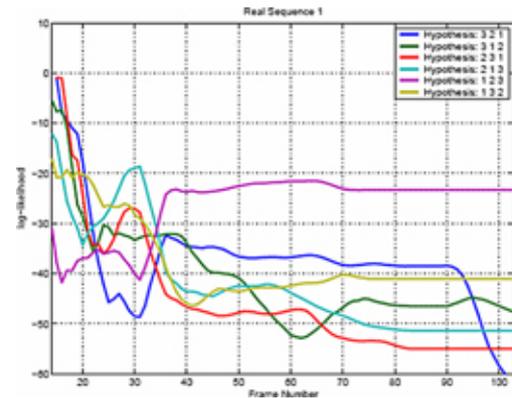


Figure 4. Variation of global correspondence hypotheses shown as a plot of log-likelihood of different hypotheses against time in UAV videos.

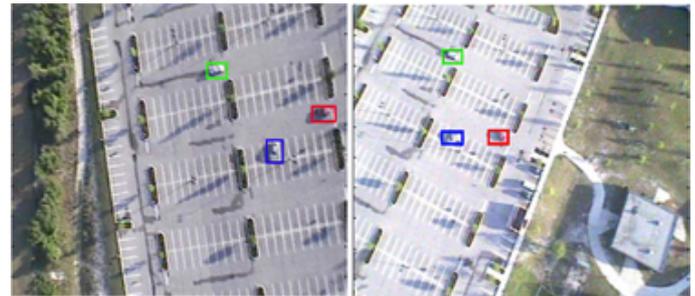


Figure 5. Sample time-synchronized UAV frames: (left) UAV1, (right) UAV2. Red, green and blue bounding boxes represent same objects across different UAVs.

between these bi-partitions are the correspondence likelihood estimates. A more complex scenario can have multiple UAVs detecting several objects simultaneously. In order to obtain a globally optimal trajectory correspondence, the solution is equivalent to finding maximum matching of the split G^* of the directed acyclic weighted graph D , as detailed elsewhere.²

Experiments were performed on UAV videos and the effectiveness of the likelihood maximization estimate for the global correspondence hypotheses is shown in Figure 4. The chart shows the results where the correct hypothesis is identified when objects show more non-collinear motion. Figure 5 shows sample frames from two different UAVs detecting the same three objects on the ground. The videos had short temporal overlap but the estimated object correspondence was correct.

We have presented an approach for the detection and tracking of objects across multiple airborne electro-optical sensors. In this approach, there is a restriction that some UAVs must have an



overlapping field of view for some part of the object trajectory. We are currently looking into ways to remove this constraint for global object correspondence across UAVs.

Author Information

Mubarak Shah, Asaad Hakeem, and Arslan Basharat

Computer Vision Lab

University of Central Florida

Orlando, FL

References

1. S. Ali and M. Shah, *COCOA – Tracking in Aerial Imagery*, **Demo: IEEE International Conference on Computer Vision**, China, October 2005.
2. Y. Sheikh and M. Shah, *Object Tracking Across Multiple Independently Moving Cameras*, **IEEE International Conference on Computer Vision**, China, October 2005.