COCOA - Tracking in Aerial Imagery

Saad Ali and Mubarak Shah

Computer Vision Lab,
University of Central Florida
Orlando, FL, USA

ABSTRACT

Unmanned Aerial Vehicles (UAVs) are becoming a core intelligence asset for reconnaissance, surveillance and target tracking in urban and battlefield settings. In order to achieve the goal of automated tracking of objects in UAV videos we have developed a system called COCOA. It processes the video stream through number of stages. At first stage platform motion compensation is performed. Moving object detection is performed to detect the regions of interest from which object contours are extracted by performing a level set based segmentation. Finally blob based tracking is performed for each detected object. Global tracks are generated which are used for higher level processing. COCOA is customizable to different sensor resolutions and is capable of tracking targets as small as 100 pixels. It works seamlessly for both visible and thermal imaging modes. The system is implemented in Matlab and works in a batch mode.

Keywords: UAV, Tracking, Aerial Imagery, Detection

1. INTRODUCTION

In the current application we are concerned with the tracking of multiple moving objects in videos taken from an aerial platform. In order to extract any useful information about the motion of the objects we need to detect and track them over long durations. This is a challenging task as target sizes are small and they must be acquired and tracked through changing terrain, evolving appearance and frequent occlusions. In addition, objects can enter and leave the field of view of the camera at arbitrary points. Unmanned air vehicles often utilize visible and thermal imaging modes. Therefore, an efficient indexing system will also have to work seamlessly across these modes. We propose an efficiently engineered system which reliably locates and tracks object of interests under these settings by using a modular approach. The approach involves three modules which include platform motion compensation, motion detection and object tracking. Figure 1 shows the block diagram of the system where it receives the video and the telemetry information from the airborne sensor. After the reception of the video motion compensation module registers the video and generates corresponding aligned images. In order to increase the flexibility of the system multiple video registration methods are integrated into COCOA. The list includes gradient based method, feature based method and telemetry based method. Second module is comprised of moving object detection which takes the registered images and detects moving objects from them. At this stage accumulative frame differencing and background modelling are used to figure out the background and foreground pixels. The detected regions are further processed by using a level set based segmentation approach to acquire the contours of the objects. Contours help the system in modelling the exact appearance and shape of the target objects. Finally the regions are tracked across the video by using a blob based tracking approach.

Over the years vision researchers have proposed numerous algorithms for vision based processing of the videos taken by the aerial vehicles. Methods proposed in Mann et.al,[1] Hartley et. al[2] and Sawhney at. al[3] are popular for video alignment and mosaicing. Kumar et al.,[4] proposed a method to perform geo-registration by utilizing a database of previously geo-registered images and digital elevation map (DEM). For indexing of UAV videos Irani et al.,[5] proposed to convert video stream into a series of mosaics that provide complete coverage of the original video. The video mosaics are than used as visual indices into a video database for analysis, storage and retrieval of video footage. Roth[6] proposed another approach for organizing and searching the UAV video database by

Further author information: (Send correspondence to Dr. Mubarak Shah, Saad Ali)
Dr. Mubarak Shah.: E-mail: shah@cs.ucf.edu
Saad Ali.: E-mail: sali@cs.ucf.edu

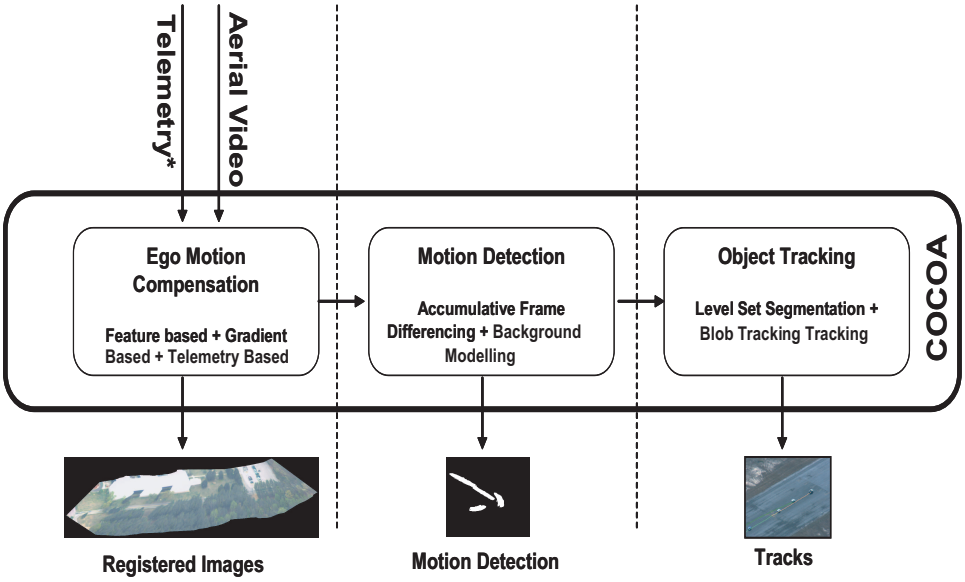

Figure 1. Components of our aerial video tracking system.

employing an indexing structure comprising of interest point descriptors of key frames. The breakdown of the research in this area shows that considerable amount of effort is being invested in developing techniques that can solve some specific aspect of UAV video processing. Very little attention is being paid to building complete systems that can automate the whole video exploitation pipeline. COCOA is an effort in the direction of coming up with systems that streamline the task of UAV video processing for the purpose of generating high value intelligence information. Now we will discuss the details of different modules of COCOA.

The paper is organized as follows. Section 2 will discuss motion compensation, moving object detection and object tracking modules of COCOA. Section 3 will present some results on different sequences.

2. TRACKING SYSTEM

Tracking system is comprised of three modules: Ego Motion Compensation, Independent Motion Detection and Object Acquisition and Tracking.

2.1. Ego Motion Compensation

Ego motion compensation module of COCOA accounts for the continuous motion of the camera mounted on the aerial platform. Motion compensation contributes to the working of the system in two ways. Firstly, it helps in detecting independently moving ground targets as after compensation in any two neighboring frames the intensity of only those pixels will be changing that belong to the moving object. Secondly, by registering the whole video with respect to one global reference we are able to get a meaningful representation of the object trajectories which can be used to describe the entire video. It uses a combination of feature-based method, gradient-based method and platform meta data for frame to frame registration. This helps in exploiting both the robustness of feature-based methods and the accuracy of gradient-based methods. The feature based approach is similar to the one proposed by Hartley et. al.[2] In this approach, the Harris corner detector is used to detect points in the frames to be matched, followed by coarse correspondence using a simple correlation matcher. The tentative matches are then filtered using RANSAC to find the correspondences that best fit a homography. A gradient based alignment, proposed by Mann et. al[1] was then used to refine the estimate of the inter-frame homography. Since independently moving objects are expected in our scenario, we used iteratively re-weighted least squares to solve the over constrained linear system of equations robustly. When telemetry is available for the moving platform, it can be used to initialize the registration algorithm. In COCOA, an additional feature is provided

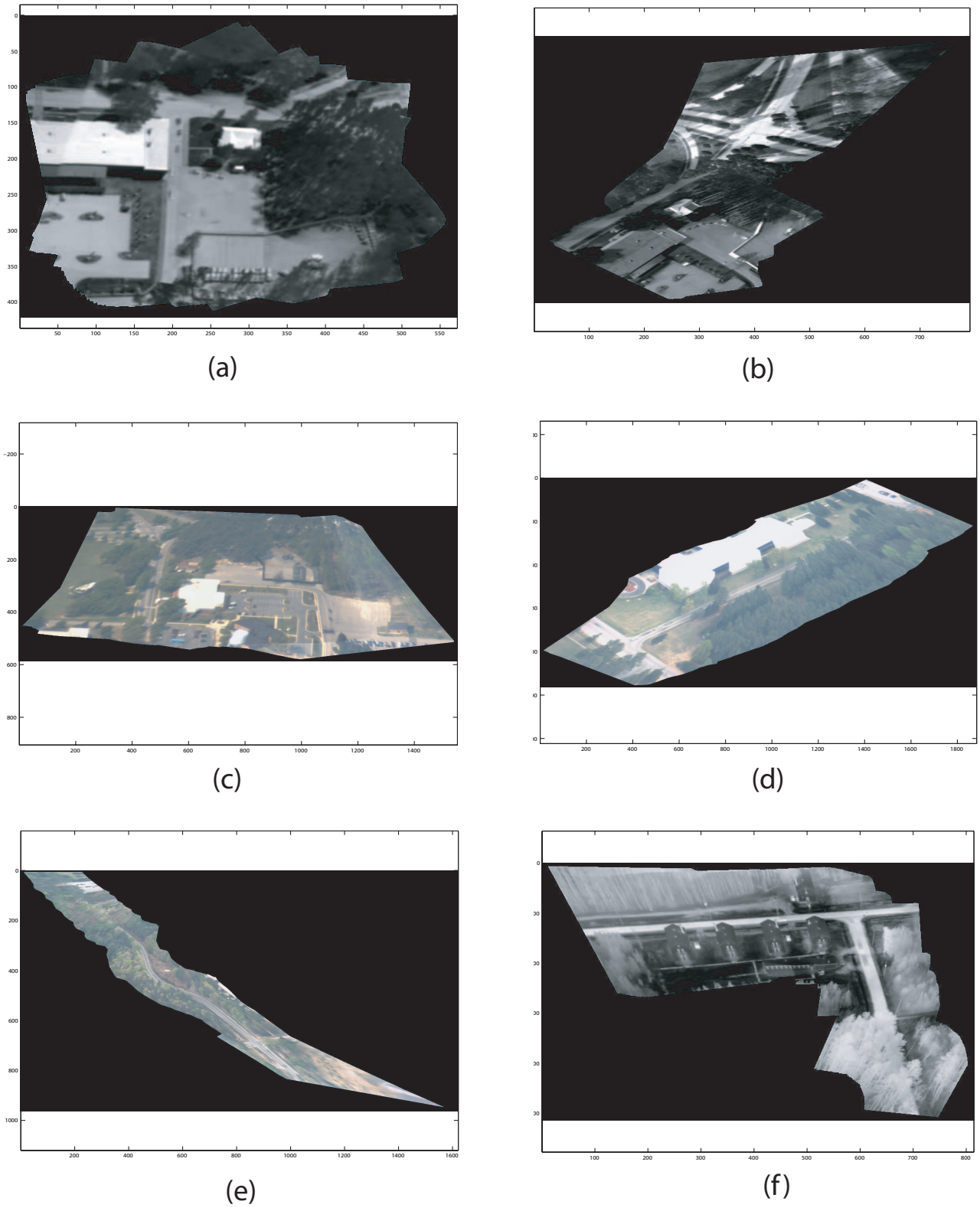

Figure 2. Global mosaic of six different sequences from the VIVID-2 data set. Telemetry information was used to initialize the estimate of frame to frame translation parameters followed by the use of gradient-based registration to obtain the full projective parameters of the transformation.

for registering videos in small blocks. This helps in avoiding the accumulation of errors. Figure 1 shows mosaics generated from the videos available in DARPA VIVID data set.

2.2. Moving Object Detection

To detect motion of independently moving objects, such as cars, trucks, people or motorbikes, we incorporated two methods into COCOA which are *Accumulative Frame Differencing* and *Background Subtraction*. In addition, to accommodate sequences of hundreds of frames, we performed differencing with respect to a sliding reference coordinate. After computing the difference of each frame with respect to it's p neighboring frames the evidence is accumulated by adding the difference images to obtain a sum image. The log-evidence at each frame is histogrammed. The large peak corresponds to background pixels and the smaller peak corresponds to the foreground pixels. Figure 2.2 shows the flow of this process for a sequence of images.

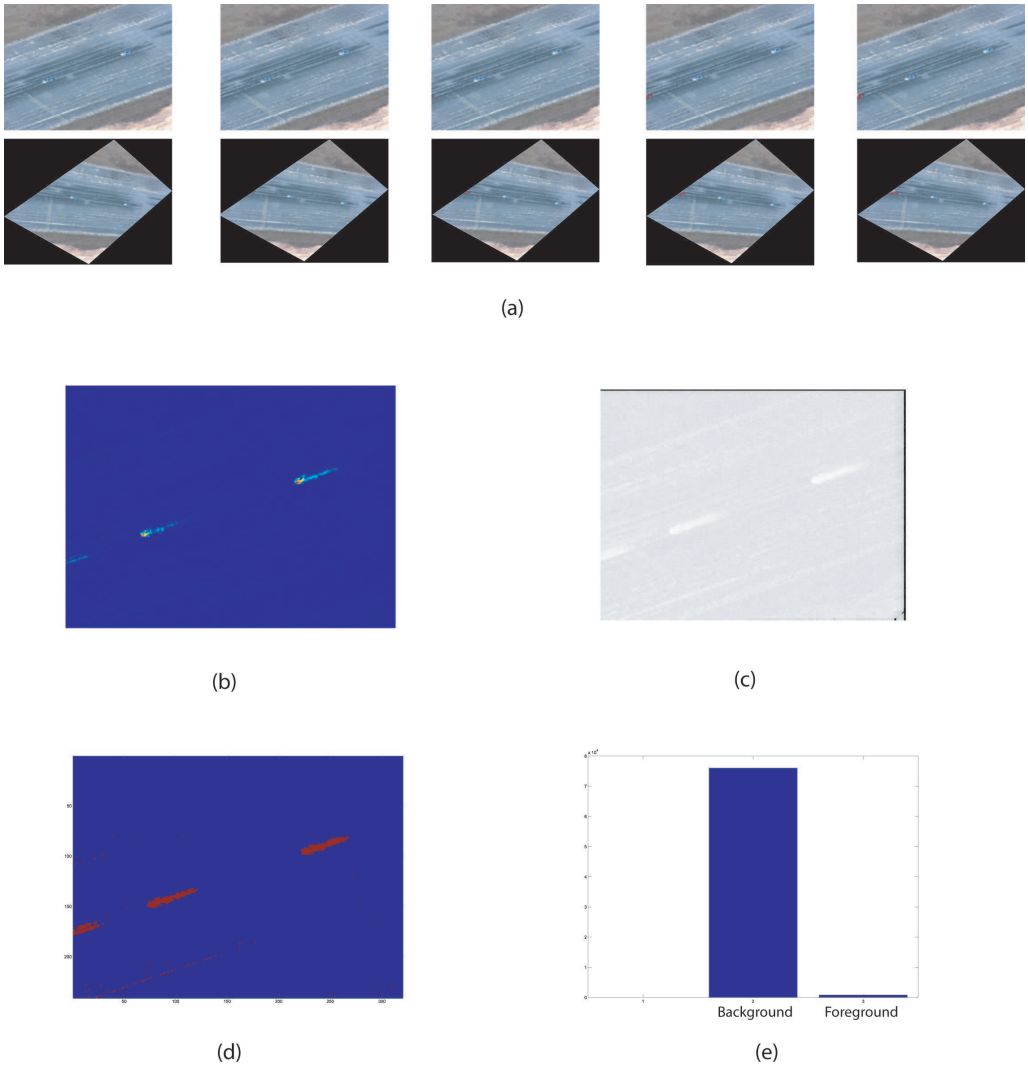

Figure 3. (a) Shows a sequence of frames with their corresponding aligned images. In this sequence we are interested in finding out the foreground objects present in the center frame. (b) Shows image obtained by take the difference of pixel intensities between the first and the third aligned image. (c) Sum of the four difference images of this sequence. (d) Shows log of the sum image from (c). Three moving foreground objects are detected in the image. (e) Histogram of the log evidence. Larger peak represents all the pixel that belong to the background while smaller peak is for the foreground pixels.

Second motion detection algorithm integrated into COCOA is the background subtraction. The background subtraction is performed in a hierarchical manner consisting of two stages i.e., Pixel level and Frame level processing as proposed by Javed et. al.[9]

2.3. Tracking

Tracking is critical for obtaining meaningful tracks that capture the motion characteristics over longer durations of time. Blob tracking approach is used to perform multi-target tracking. Regions of interest, or blobs, in successive frames are given by the motion detection module. Each blob is represented by its own appearance and shape models. Temporal relationship between the blobs are established by computing appearance and shape similarity score. The relationship is established if the score is above a threshold. Figure 4 shows a sequence in which three cars are being tracked by using the blob tracking algorithm.

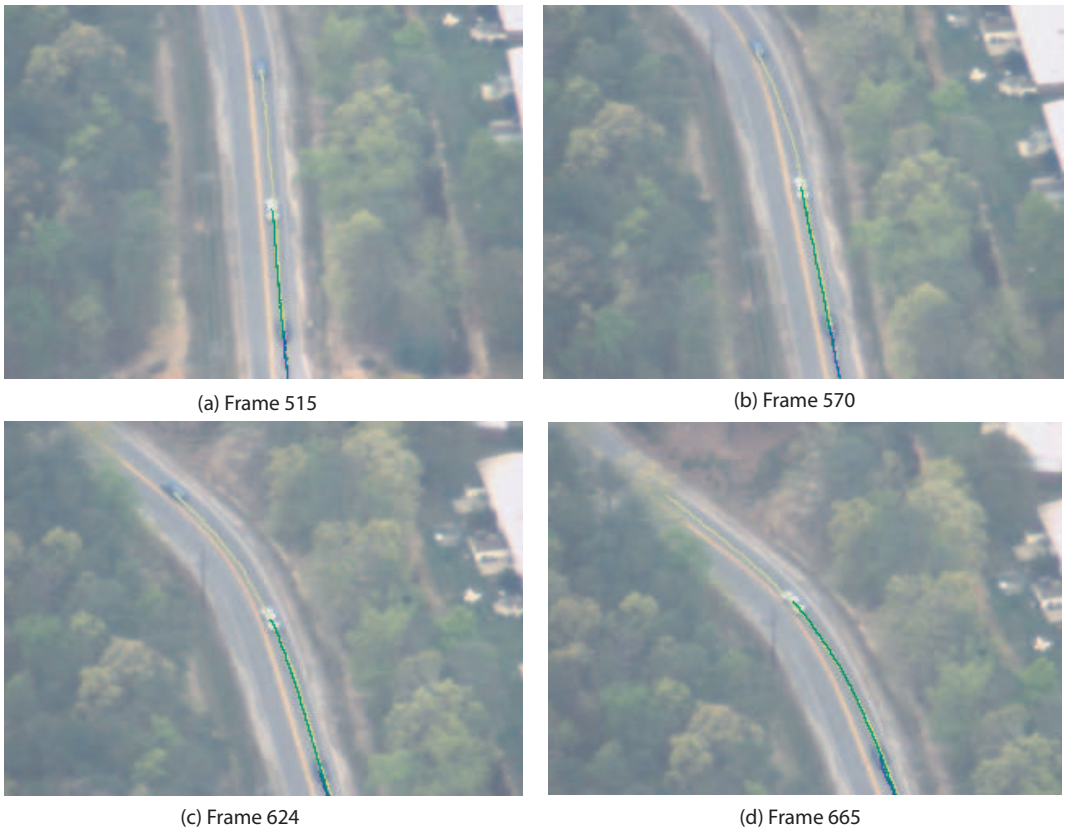

Figure 4. Tracking the target object using COCOA. First row shows the tracking result while the second row shows the tracks generated by each target object. Tracking is done in global frame of reference, however the tracks are superimposed on the local frame of reference in this case.

Occlusions are also handled by COCOA by using method proposed by Javed.[11]

3. RESULTS

Object detection and tracking modules of COCOA were evaluated using the VACE evaluation process. In this evaluation process the task was defined as detection and tracking of moving vehicles when a substantial part of the vehicle body is visible. Oriented bounding boxes around the vehicle were required to be annotated. Evaluation data set consisted of 50 sequences (each 1 minute long) from VIVID-2 corpus. It included 17 long range, 17 short range and 16 infrared sequences.

REFERENCES

1. S. Mann and R. Picard, "Video Orbits of the Projective Group: A Simple Approach to Featureless Estimation of Parameters," *IEEE Transactions on Image Processing*, September 1997, pp. 1281-1295.
2. R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2000.
3. H. S. Sawhney, S. Hsu and R. Kumar, "Robust Video Mosaicing Through Topology Inference and Local to Global Alignment", *Proceedings of the 5th European Conference on Computer Vision-Volume II*, June 1998, pp. 103-119.
4. R. Kumar, H.S. Sawhney, J.C. Asmuth, A. Pope, and S. Hsu, "Registration of Video to GeoReferenced Imagery," *International Conference on Pattern Recognition*, August 1998, pp. 1393-1400.
5. M. Irani and P. Anandan, "Video Indexing Based on Mosaic Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1998, pp. 905-921.
6. G. Roth and W.R. Scott, "An Image Search System for UAVs," *UVS Canada's Advanced Innovation and Partnership Conference*, November 2005.
7. G. Medioni, I. Cohen, F. Bremond, S. Hongeng and R. Nevatia, "Event Detection and Analysis from Video Streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2001, pp. 873-889.
8. A. Yilmaz, X. Li and M. Shah, "Contour Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, November 2004, pp. 1531-1536.
9. O. Javed , K. Shafique and M. Shah, "A Hierarchical Approach to Robust Background Subtraction Using Color and Gradient Information," IEEE Workshop on Motion and Video Computing, Orlando, December 2002.
10. A. Yilmaz, K. Shafique, T. Olson, N. Lobo and M. Shah, "Target Tracking in FLIR Imagery Using Mean-Shift and Global Motion Compensation," IEEE Workshop on Computer Vision Beyond Visible Spectrum, December 2001.
11. O. Javed, "Thesis: Scene Monitoring with a Forest of Cooperative Sensors", University of Central Florida, Orlando, 2005.