

Outdoor Environment Monitoring with Unmanned Aerial Vehicles

Claudio Piciarelli, Christian Micheloni, Niki Martinel,
Marco Vernier, and Gian Luca Foresti

Department of Mathematics and Computer Science, University of Udine
Via delle Scienze 206, 33100 Udine, Italy

{claudio.piciarelli,christian.micheloni,
niki.martinel,marco.vernier,gianluca.foresti}@uniud.it

Abstract. This work addresses the problem of video surveillance of outdoor environments with unmanned aerial vehicles (UAV). Specifically it proposes a two-step approach, with an initial offline stage in which a mosaic of the zone to be monitored is built from video sequences. The second step tackles with the problem of online detection of relevant differences between the acquired images and the mosaic model. A GPS-assisted approach is proposed to deal with efficiency issues in this online step. Experimental results prove that the proposed approach can be used to detect relevant changes in the specific case of road safety assurance in dangerous zones.

Keywords: UAV, aerial imaging, change detection.

1 Introduction

This work aims to explore the monitoring capabilities of a system composed by a video sensor mounted on an unmanned aerial vehicle (UAV), and specifically on a multicopter drone. Unmanned aerial surveillance has multiple advantages both in civil and military applications. In particular these systems are useful when the monitoring activity is particularly tedious (constant patrolling in large environments) or dangerous because of environmental conditions (e.g. observing a nuclear incident site) or because of security issues (e.g. monitoring hostile territories in peacekeeping missions).

UAV-based computer vision is a field of research that has been gathered an increasing attention during the last years. For example the WITAS project [13], developed from 1999 to 2005, aimed to study how UAV-based systems could be exploited in different scenarios (traffic monitoring, surveillance, emergency systems, photogrammetry, etc). Other examples are the systems by Nikolos et al. for fire detection [11] and multiple UAV coordination [12], or the work by Li et al. for monitoring areas hit by earthquakes [8].

More in general, UAV-based systems have been used for two main aims: imaging and object detection. In the first case the goal is to exploit the UAV in order to create large views of specific zones (mosaicking) or improve other aerials views

with finer details. This is the case of the work by Fan et al. [3] that enhances satellite images with aerial views with better resolution. Lin and Medioni [9] propose an image mosaicking technique to reduce the distortions arising when registering aerial images. Regarding object detection and tracking, one of the first works is by Cohen and Medioni [10], where a graph-based dynamic model allows to keep track of detected moving objects in the scene. The moving objects are detected by optical flow analysis. Gaszczak et al. [5] detect moving people by using both visible and thermal cameras. Jackson and Goshtasby [7] rely on image registration to compensate the camera motion and detect image differences caused by moving objects in the scene.

In this paper we propose a UAV-based system to detect relevant changes in outdoor environments. The proposed method relies on a GPS-assisted and feature-based approach to speed up the frame-to-background matching phase.

2 System Architecture

The logical architecture of the system is split in two main parts, the mosaic building module and the change detection module. The mosaic building tasks are performed offline on video sequences previously acquired by the UAV flying over the area to be monitored. The system extracts relevant features from each frame of the video sequence and performs feature matching in order to properly identify the spatial relations between images. This information is used to stitch the processed frames into a mosaic, i.e. a single large image representing the entire area to be monitored. Proper blending techniques are adopted in order to avoid image artifacts caused by camera exposure differences. Finally, a set of auxiliary information is stored in order to facilitate the subsequent online step. The information consists in a set of image features and their locations extracted from the entire mosaic, and a set of GPS-measured real-world coordinates associated to different points of the mosaic.

The change detection module consists in a sequence of tasks performed online in order to detect relevant changes between the currently acquired images and the previously built background model (the mosaic). The images and the mosaic are first roughly registered using GPS data, then the registration is further refined by matching the features extracted from the current image and the ones computed offline from the mosaic. Finally, the registered images and the mosaic are compared in order to extract the zones where changes have occurred. Figure 1 summarizes the logical architecture of the system.

Regarding the physical architecture, the experiments have been performed using a home-built, remotely controlled hexacopter. The UAV is equipped with a color full-HD camera streaming the video signal over a wireless connection to the main processing unit. The camera has a built-in image stabilization system reducing the effects of UAV vibrations. An IMU (Inertial Measurement Unit) is used to keep the camera's optical axis perpendicular to the ground plane, in order to acquire top-view images of the observed area. The hexacopter also

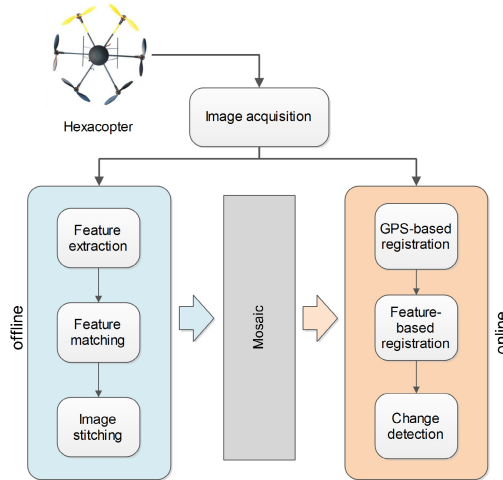


Fig. 1. Logical architecture of the proposed system

mounts a GPS system with magnetic compass in order to tag each frame with position and orientation information. The UAV flies at constant height from the ground.

3 Mosaic Building Module

The mosaic building module aims to build a large image of the area to be monitored from a video sequence acquired by the UAV. The processing is performed offline, since the mosaic is meant to be used only as a reference image for later (online) comparisons.

The UAV mounts a gyroscope-controlled camera which actively keeps the optical axis perpendicular to the ground plane; moreover we hypothesize that the height of the observed objects is negligible with respect to the UAV height of flight. Under these assumptions different images acquired by the system could be registered by means of affine transformations. However we choose to model these transformations as homographies in order to avoid misalignments, e.g. due to small errors in the gyroscope-based camera orientation control system.

The first step of the mosaic building module thus consists in detecting robust image features that can be used for alignment purposes. For this scope, we adopted the widely used SURF features [1]. Robust feature matching between two different images is found using RANSAC: a small subset of the matching features is randomly chosen, then the corresponding homography is computed and the projection error on the remaining features is measured. The process is iterated several times and the homography minimizing the error is chosen.

Since consecutive frames largely overlap, not all of them are really useful to build the final mosaic. For this reason we adopted a dynamic frame selection

algorithm in order to identify the *key frames*, i.e. the frames that will compose the mosaic. We start with an arbitrarily chosen dropping rate d , meaning that only one frame every d is a key frame. Features are extracted from the current key frame and, if a correspondence in the previous key frame is found for less than 30% of the total features, the dropping rate is decreased. Conversely, if the ratio between matched features and total number of features is above 70% the dropping rate is increased. Since the processing is performed offline, the dropping rate can be freely modified until a proper value is found.

Using the above described procedure, homographies between consecutive key frames can be computed. However, two non-consecutive key frames could still overlap, e.g. if the UAV follows a circular path. In this case the system could incur in poor image alignment due to accumulating errors in each image registration step. The system tries to detect these cases by using the GPS data and searching for feature matches between non-consecutive key frames lying within a fixed range. If a match is found, a global optimization technique must be adopted in order to find a set of proper image transformations that minimizes the overall registration error. For this purpose, we adopted the bundle adjustment technique described in [2].

Finally, gain compensation and image blending are needed to merge the chosen key frames in a single mosaic avoiding image artifacts on the frame borders. We again used the techniques proposed in [2]. For each pair of registered images I_1, I_2 , gain compensation is expressed as a minimization problem that searches the gain factors g_1, g_2 that minimize the total error $\sum_{(x,y) \in \mathcal{R}} (g_1 I_1(x,y) - g_2 I_2(x,y))^2$ computed on the overlapping region \mathcal{R} . Blending is obtained by associating linearly decreasing weights to image pixels, starting from image center (maximum weight) to image borders.

3.1 Offline Pre-computations

The output of the aforementioned procedure is a large mosaic image, obtained by stitching several video frames together, representing a top-view of the entire outdoor area to be monitored. This mosaic will be used as a reference image in later online processing to detect relevant changes. In order to match the online images with the correct portion of the mosaic, both GPS data and feature matching will be used. Some of the required data can be pre-computed offline to speed up the online processing.

Regarding GPS data, the center of each stitched image in the mosaic is labeled with GPS coordinates and orientation. The assumption that the center of the image has the same latitude/longitude coordinates measured by the GPS system on the UAV holds since the camera optical axis is kept perpendicular to the ground floor.

For each image i , the orientation is stored as a unitary vector (x_i^N, y_i^N) denoting the North direction. In this case we compute the angle $\theta_i = \arctan(y_i^N/x_i^N) - \pi/2$ for each image. The entire mosaic is then rotated by $\theta = 1/n \sum_{i=1}^n \theta_i$ (where n is the total number of stitched images) in order to align its y axis with the North direction.

After rotation, we compose the two sets $S_x = \{(x_i^M, x_i^G)\}_{i=1}^n$ and $S_y = \{(y_i^M, y_i^G)\}_{i=1}^n$, where (x_i^G, y_i^G) are the GPS coordinates of point i and (x_i^M, y_i^M) are its pixel coordinates in the mosaic image. Assuming that the spherical latitude/longitude coordinates system can be locally approximated as a Cartesian system, we apply a robust linear regression to both sets independently, in order to estimate two functions $F_x(\cdot), F_y(\cdot)$ converting pixel coordinates into world coordinates. The two functions are easily invertible, thus it is also possible to estimate the image coordinates from world coordinates by computing $F_x^{-1}(\cdot)$ and $F_y^{-1}(\cdot)$. This property will be later used for GPS-based rough image registration.

The coordinate conversion functions have the form $F_x(x^M) = \alpha x^M + c$ (similarly for the y coordinates) where α and c could be non-robustly estimated by solving the system $A(\alpha, c)^T = b$ with the Least Squares algorithm, where

$$A = \begin{pmatrix} x_1^M & 1 \\ \vdots & \vdots \\ x_n^M & 1 \end{pmatrix} \quad b = \begin{pmatrix} x_1^G \\ \vdots \\ x_n^G \end{pmatrix} \quad (1)$$

In order to get rid of GPS inaccuracies we achieved robust estimation by using the Iterately Reweighted Least Squares algorithm [6], and thus iteratively computing at each step k the following equations:

$$\begin{aligned} x^k &= (A^T W^k A)^{-1} A^T W^k b \\ W^k &= \text{diag}(w_1^k, \dots, w_n^k) \end{aligned} \quad (2)$$

where the weights w^k are a nonlinear function of the residuals. In particular, defining $r = A(\alpha, c)^T - b$, using the Lorentzian estimator we have $w_i^k = 1/(r_i^2 + \eta)$, with η constant term.

Finally, the entire mosaic is searched for SURF image features. Both the feature positions and the feature descriptors are stored for later retrieval. This way the online feature-based image alignment process will require the extraction of features only on the newly acquired images, since the mosaic is static and its features do not change.

4 Change Detection Module

The online component of the system aims at detecting relevant changes in the environment by comparing the acquired images with the mosaic. In order to have a rough estimate of the location of the current frame within the mosaic, the GPS information is used. Given the measured GPS coordinates (x^G, y^G) , the corresponding position on the map (x^M, y^M) is given by:

$$x^M = F_x^{-1}(x^G) \quad y^M = F_y^{-1}(y^G) \quad (3)$$

where $F_x(\cdot), F_y(\cdot)$ are the functions described in section 3.1.

Then, SURF features are extracted from the current frame. Feature correspondences are searched with the mosaic features within a window centered in

(x^M, y^M) and a radius equal to the width of the current image. Since the SURF features are rotationally invariant, the match can be found even if the two images do not have the same orientation. Outliers are rejected using RANSAC and the remaining feature matches are used to compute the homography H used to register the frame with the mosaic. If (x, y) are pixel coordinates in the image frame, the corresponding point in the mosaic is thus $(x'/\lambda, y'/\lambda)$, where $(x', y', \lambda)^T = H(x, y, 1)^T$.

Once the registering homography has been computed, the difference image D is computed by image subtraction between the current frame I and the mosaic M :

$$D(x, y) = \begin{cases} \|M(x'/\lambda, y'/\lambda) - I(x, y)\|_2 & \text{if } (x'/\lambda, y'/\lambda) \in M \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The L^2 norm is required since each pixel value is a 3-elements vector representing the red, green and blue components of the pixel.

The difference image is finally thresholded in order to highlight only relevant changes. In order to avoid the use of arbitrary fixed thresholds, we adopted an automatic thresholding technique based on fast computation of Euler numbers [4]. The result is a binary images where connected components are considered regions of relevant changes. Smaller regions are discarded as noise, the remaining ones are labeled with their world coordinates and signaled to a human operator.

5 Experimental Results

The proposed work was primarily inspired by the need to guarantee the security of peacekeeping missions in dangerous zones. In these contexts, often civil or military vehicles are required to move across potentially unsafe areas, where they can be the target of attacks by terrorists or by guerrillas or commando forces in unconventional warfare scenarios. These attacks are often conducted by using the so-called “roadside bombs”, i.e. improvised explosive devices (IED) hidden near road borders. For example, it is estimated that over 66% of the US-led coalition deaths in the Afghanistan war have been caused by this type of devices.

In order to simulate this kind of scenario, we acquired video sequences of a country road during different days. Figure 2 shows some frames of a video sequence. The frames have been offline processed in order to build the final mosaic using the techniques described in section 3. The dropping rate d has been automatically computed by the system and assumed an average value of 18, meaning that on average only one frame every 18 has been processed for mosaic creation. SURF features have been detected in the selected frames and have been used to compute the required image registration transforms. Images are then warped and blended in order to create the final mosaic, shown in Figure 2(d). The mosaic does not show relevant errors, since all the selected frames have been correctly stitched.

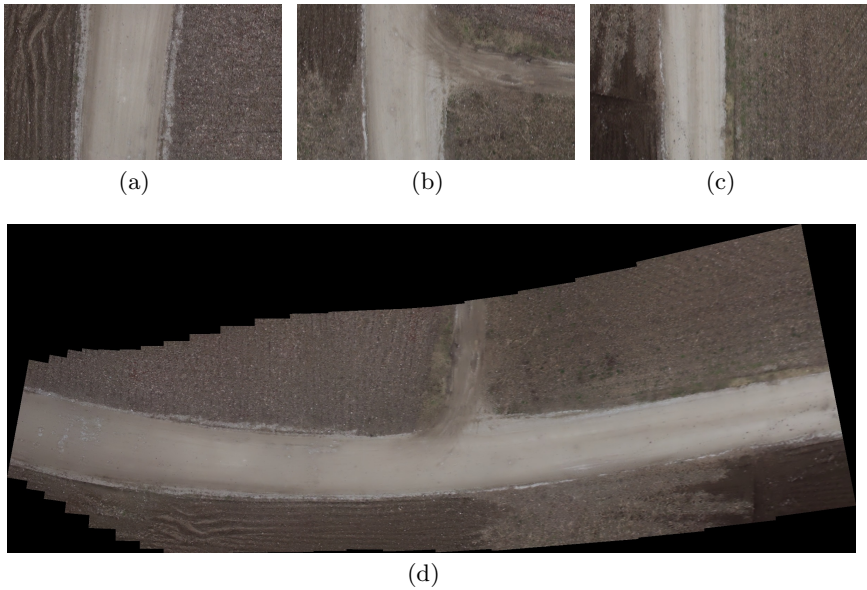


Fig. 2. (a)-(c): Some frames from the acquired video sequence; (d) the corresponding mosaic

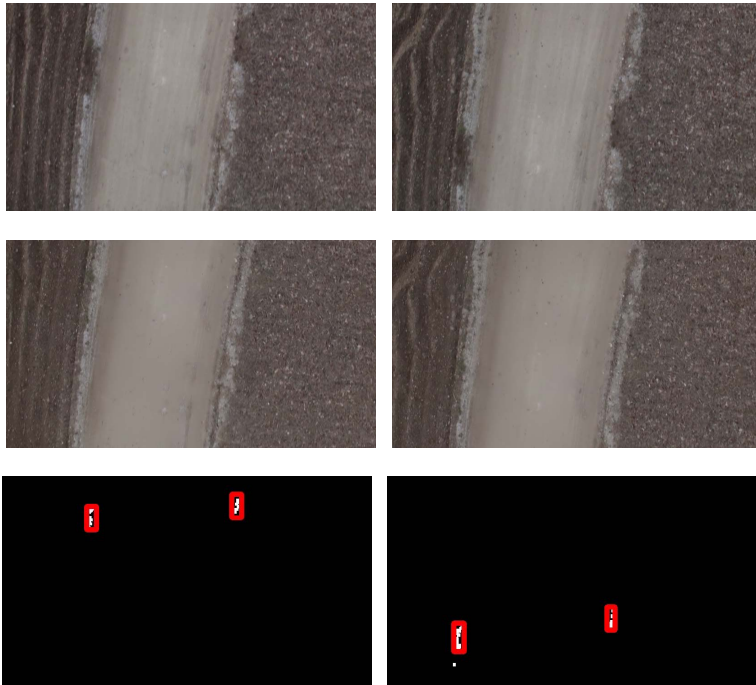


Fig. 3. Detection results. First row: acquired frames; second row: the corresponding mosaic portions; third row: the detected zones

Another video sequence of the same zone has been acquired 20 days later under similar lighting conditions, after digging two holes near the road borders. The system has been run in online detection mode, by using the previously computed mosaic as a background model. After thresholding, the difference images have been filtered in order to drop small connected components (less than 40 pixels) in order to reduce noise. As a result, the two holes have been correctly detected, as shown in figure 3. No false detections occurred in the rest of the video sequence. The test has been repeated 10 times, always with the same success rate.

6 Conclusions

With this work we have investigated the possibility of using UAV-based vision systems to monitor outdoor areas for relevant change detection. The system works on a two steps basis: in the first step a background mosaic is built by stitching together several frames of a video sequence. In the second online step the frames of a new video sequence are automatically matched with the correct portions of the mosaic. The differences between the two images allows to highlight the zones where relevant changes have occurred. Preliminary experimental results have shown good performances. Future developments will focus on robust change detection for outdoor environments, considering in particular the negative effects of different lighting conditions.

References

1. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
2. Brown, M., Lowe, D.: Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision* 74(1), 59–73 (2007)
3. Fan, B., Du, Y., Zhu, L., Tang, Y.: The registration of UAV down-looking aerial images to satellite images with image entropy and edges. In: *3rd International Conference on Robotics and Applications*, Shanghai, China (2010)
4. Foresti, G.L., Micheloni, C., Snidaro, L., Remagnino, P., Ellis, T.: Active Video-based Surveillance Systems: the low-level image and video processing techniques needed for implementation. *IEEE Signal Processing Magazine* 22(2), 25–37 (2005)
5. Gaszczak, A., Breckon, T.P., Han, J.W.: Real-time People and Vehicle Detection from UAV Imagery. In: *Proc. SPIE Conference Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, vol. 7878, No. 78780B (2011)
6. Huber, P.J.: *Robust statistics*. John Wiley & Sons, Inc. (2005)
7. Jackson, B., Goshtasby, A.: Registering video images using the projective constraint. *IEEE Trans. Image Processing* 19(3), 795–804 (2010)
8. Li, C., Zhang, G., Lei, T., Gong, A.: Quick image-processing method of UAV without control points data in earthquake disaster area. *Transactions of Nonferrous Metals Society of China* 21(3), 523–228 (2011)
9. Lin, Y., Medioni, G.: Map-Enhanced UAV Image Sequence Registration and Synchronization of Multiple Image Sequences. In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (2007)

10. Cohen, I., Medioni, G.: Detecting and tracking moving objects for video surveillance. In: IEEE Proc. Computer Vision and Pattern Recognition, Fort Collins, CO, USA (1999)
11. Nikolos, I.K., Tsourveloudis, N.C., Valavanis, K.P.: A UAV vision system for airborne surveillance. In: Proceedings of the IEEE International Conference on Robotics and Automation, New Orleans, LA, USA, pp. 77–83 (2004)
12. Nikolos, I.K., Tsourveloudis, N.C., Valavanis, K.P.: Multi-UAV experiments: Application to forest in Multiple Heterogeneous Unmanned Aerial Vehicles. Springer Tracts in Advanced Robotics, pp. 207–228. Springer, Berlin (2007)
13. Wzorek, M., Doherty, P.: The WITAS UAV Ground System Interface with Focus on Motion and Task Planning. In: System Demonstration Abstracts of the 16th International Conference on Automated Planning and Scheduling (2006)