2016 IEEE International Conference on Systems, Man, and Cybernetics • SMC 2016 | October 9-12, 2016 • Budapest, Hungary

Stable aerial image registration for people detection from a low-altitude aerial vehicle

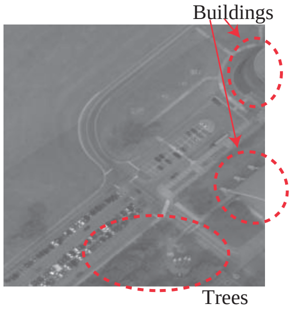
Yumi Iwashita † , Yuki Takefuji †, Ryo Kurazume †
†Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
Email: yumi@ieee.org

***Abstract*—In this paper, for the purpose of monitoring people moving on the ground from a low-altitude aerial vehicle, we propose a method for stable image stabilization using planar information of the ground. In existing methods the homography-based method has been popularly used for image stabilization. However, in case that captured images consist of non-planar areas, the performance of the homography-based method decreases. The essential-based method works well in this scene, but the accuracy gets worse in case that images mainly include planar areas. Thus in the proposed method, we utilize the Geometric Robust Information Criterion (GRIC) model to choose a method from the homography-based method or the essential-based method. After the selection of the method, we extract a planar area in the scene by fitting a plane using RANSAC, followed by detection of people on the ground with high accuracy. Experimental results confirm that the effectiveness of the proposed method.**

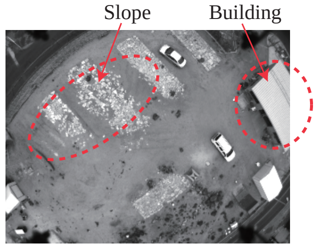
I. Introduction

Unmanned aerial vehicles, which are also called as drones, have received great attention for various purposes, such as security systems from the sky, pollution monitoring, and also recreational use. We have focused on developing a system for tracking people from the sky for security purpose [8] [7], and generally a video stabilization technique is applied to aerial images before applying the process of the people tracking. There are several techniques for video stabilization, for example, a homography-based method [2] [10], and a method based on essential matrix estimation [11] [6] [3] (assuming that intrinsic camera parameters are known). The homography-based method works well if the scene in the video consists mostly of flat regions. In case that videos are recorded in higher altitude flight as shown in Fig. 1 (a), the non-planar areas with buildings, trees and hills could be ignored due to the enough height of the airship compared with the objects on the ground. For such a situation the homography-based method worked well for video stabilization. The essential matrix-based method was mainly used for relative pose estimation between images, and the image registration was done after the estimation of a scale factor [13]. An important issue to be considered regarding the essential matrix is that the accuracy of the pose estimation gets worse in certain conditions, such that the scene is a planar area and there are noises on detected points on images [3].

This research was performed when the first author was at Kyushu University. The first author is now at Jet Propulsion Laboratory, California Institute of Technology. This paper does not contain any results from the researches at Jet Propulsion Laboratory.



(a)



(b)

Fig. 1. (a) An example of aerial images captured at a high-altitude platform, (b) an example of aerial images captured at a low-altitude platform. In general a homography-based method works well with images taken at high altitude (i.e. (a)). On the other hand the homography-based method sometimes has difficulties with images taken at low altitude, and an essential-based method works well.

In case that an unmanned aerial vehicle flies at low altitude over an area including non-planar regions (Fig. 1 (b)), the homography-based method loses stabilization accuracy with respect to the ground. On the other hand, the method based on the essential matrix works well in this scene. However, as we mentioned, the performance of the essential-based method gets worse if the scene is a planar area, due to the degeneracy of eigenvalues of the matrix. Thus the method which can choose a proper method to the scene, either the homography-based method or the essential-based method, is necessary.

In this paper, for the purpose of monitoring people moving on the ground from a low-altitude aerial vehicle, we propose a method for stable image stabilization using planar information of the ground. In this method, we utilize the Geometric Robust Information Criterion (GRIC) model to choose a method from the homography-based method or the essential-based method. After the selection of the method, to detect people on the ground with high accuracy, we obtain 3D point clouds using relative pose information, which is derived from a matrix from the selected method, between two images, and then we extract a planar area in the scene by fitting a plane using RANSAC (random sample consensus). Next, the images are stabilized with the homography-based method, using points on the images corresponding to 3D points on the planar area. Finally people are detected using a background subtraction method from stabilized images. The experimental results illustrate the proposed method can detect people more stable than the existing methods.

II. Related works

For the stabilization of aerial videos taken at higher altitude, the homography-based method is one of the popular methods [2] [1] [10]. To estimate the homography matrix between two images, first feature points on the 2D images are extracted by various methods, such as a corner detector, SIFT, SURF, and so on. Then the homography matrix is estimated based on various rules, such as RANSAC. Brown et al. also proposed a homography-based method for stitching images taken on the ground [4]. Lin et al. proposed a homography-based method, which recursively eliminate points which are not on planar areas [9]. Olson et al. proposed a method to stitch images based on global similarity transactions and a homography-based method which is applied to local areas [12] However, as we mentioned in the introduction, the homography-based method assumes that the scene consists mostly of planar regions.

The essential matrix-based method works well on the non-planar areas [11] [6] [3]. There are several techniques to estimate the essential matrix, such as seven-point algorithm [5] and five point-algorithm [11]. Since the seven-point algorithm is simpler and faster than the five-point algorithm, the seven-point algorithm is frequently used. However, a disadvantage of the seven-point algorithm is that the essential matrix produces a false solution in case that the scene is a planar area [11]. The computational cost of the seven-point algorithm is high due to the expensive linear algebra decompositions and polynomial solve. Botterill et al. proposed a method to overcome this problem, by using Levenberg-Marquardt optimization [3]. Compared with the seven-point algorithm, the five-point algorithm can estimate the true solution even with a planar area. However, it is reported that the accuracy of the estimation of the essential matrix gets low, if there are noises on detected points on images [3].

III. Video stabilization

This section describes a method for aerial video stabilization, which works well in both scenes: the scene with non-planar areas and the scene with planar areas. In our method, we assume that image sequences are taken from a down-looking camera, which is mounted on a moving aerial vehicle, and intrinsic camera parameters are known.

Since images are captured with a low-altitude aerial vehicle, the stabilization accuracy of the homography-based method decreases with the presence of non-planar areas. Thus to estimate the relative pose between images, we utilize the essential matrix based on the 5-point algorithm [11]. However, as we mentioned in the introduction, if the scene is a planar area and there are noises on detected points on images, the accuracy of the pose estimation based on the essential matrix gets worse [3]. In this situation the homography-based method can work better than the essential matrix, even with the presence of noise [3] [7].

The accuracy of the detection of people on the ground depends on the stabilization accuracy. Thus in this section we propose a method to stabilize aerial images with respect to the ground with high accuracy.

A. The use of GRIC to choose a proper method for the scene

In the proposed method, at first we apply a homography-based method and an essential-based method to calculate the homography matrix and the essential matrix, respectively, between image t and image $t+1$. We obtain a rotation matrix and a translation vector between images from each of homography and essential matrices, followed by the calculation of 3D points. Next, to choose the proper matrix from both matrices to the scene, we calculate GRIC score as defied as follows.

$$\begin{aligned} GRIC_m &= \sum_i \rho(\frac{e_{i,m}^2}{\sigma^2}) + \\ &\quad n_m((D-d_m)\log 2\pi\sigma^2 + 2\log\frac{c_m}{\gamma}) \\ &\quad + k_m \log n_m, \end{aligned} \tag{1}$$

$$\rho(x) = \min\{x, T_m\} \tag{2}$$

$$\begin{aligned} T_m &= 2\log(\frac{\gamma}{1-\gamma}\cdot\frac{v}{c_m}) \\ &\quad -(D-d_m)\log 2\pi\sigma^2 \end{aligned} \tag{3}$$

Here, m shows a method, either the homogprahy-based method or the essential-based method. n_m is the number of points used of a model m, and $\mathcal{L}_{MAP,m}$ is the maximum likelihood of a model m. Parameters k_m, $D-d_m$, c_m, and v are defined as Table I. $L\times L$, Δ, and $S\times S$ are the image size, the disparity value, and the area where corresponding points between images are searched. $e_{i,m}$ is a squared sum of back-projection errors of 3D points at each image and each model. In experiments, we set $\sigma^2 = 1$ and $\gamma = 0.5$. From the definition of the GRIC score equation, it is considered that the model which has a smaller GRIC score shows better performance than the other model. Thus the model which has the smaller value is selected as a proper to the scene.

TABLE I
Parameters in GRIC.

Model m	k_m	$D-d_m$	c_m	v
E	5	1	$L\times L\times\Delta$	$L\times L\times S\times S$
H	8	2	$L\times L$	$L\times L\times S\times S$

To confirm the performance of the GRIC score, we did preliminary experiments with a simulated 3D scene. In this experiments we virtually created a planar area and a non-planar area, and we put a virtual camera which moves straight in the scene. After capturing images from the virtual camera, we applied the homography-based method and the essential-based method, followed by 3D points calculation. Table II shows the average and the standard deviation of GRIC scores,

TABLE II
AVERAGE AND THE STANDARD DEVIATION OF GRIC SCORES, FOR EACH THE HOMOGRAPHY-BASED METHOD AND THE ESSENTIAL-BASED METHOD. PLEASE NOTE THAT SMALLER GRIC SCORES MEAN BETTER EVALUATION VALUES.

	Planar area		Non-planar area	
	Average	S.D.	Average	S.D.
Homography-based method	**14698.6**	273.8	15491.8	114.5
Essential-based method	15464.2	71.9	**15016.8**	46.3

for each the homography-based method and the essential-based method. Please note that smaller GRIC scores mean better evaluation values. From these results, we can see that the homography-based method and the essential-based method are successfully selected for the planar area and the non-planar area, respectively.

B. Video stabilization using points on a planar area and people detection

After selecting the proper model, we stabilize images with respect to the ground with high accuracy as follows. First we fit a plane to the 3D points, reconstructed with the selected model, by RANSAC, and feature points on the planar area are estimated. Next, we apply the homography-based method to the points on the planar area, for the stabilization of images with respect to the planar area.

For the people detection, we used a background subtraction technique. However, if we simply apply the background subtraction method to stabilized images, we extract not only the moving objects but also areas whose are above the ground, such as buildings and trees. To avoid these false positives, we utilize 3D information of reconstructed 3D points. Areas which are not judged as points on the planar area, we eliminated these areas from detected results.

IV. EXPERIMENTS

In this section, we implement and evaluate the proposed method with videos taken from a down-looking camera. To confirm the effectiveness of the proposed method, we set a camera as shown in Fig. 2 and we moved manually. In the scene we also moved a target object which moved randomly. Experiments with actual aerial images are left as future work.

A. Image stabilization with respect to the ground

We captured 20 images and Fig. 3 shows example captured images with the camera. As we can see that initially the camera moved above relatively planar area, but gradually it moved into non-planar area. Thus intuitively the homography-based method may work better in the initial frames, and the essential-based method may work better the rest of images. Table III shows GRIC scores for each the homography-based method and the essential-based method. These results show that the correct model was successfully chosen at each frame.

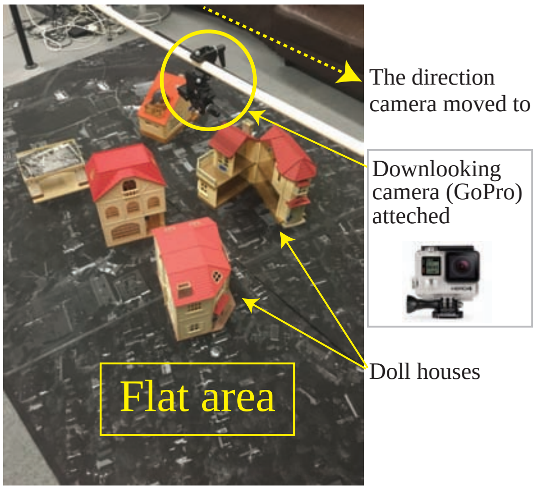


Fig. 2. Experimental settings. We used a down-looking camera, which moved manually, and 3D small houses are randomly located.

Figure 4 shows stabilized images by the proposed method, and in the figure red circles show corners of houses on the ground. We also compared with the homography-based method as shown in Fig. 5. Figure 6 shows enlarged images from Figs. 4 and 5. From these results, it is clear that the proposed method can successfully stabilize images with high accuracy compared with the homography-based method.

B. Object detection

We detected moving objects on the ground by the proposed method and the homography-based method. Figures 7 and 8 show results of detected objects, which are shown with red lines, by the proposed method and the homography-based method, respectively. Although the homography-based method has lots of false positives, the proposed method produced much better results.

V. CONCLUSION

We proposed a new method for stable image stabilization using planar information of the ground and a detection method of moving objects with less false positives. The proposed method adaptively select a proper pose estimation method, either the homography-based method and the essential matrix-based method. We confirmed the effectiveness of the proposed method with experiments.

In this paper we assumed that the camera moves continuously and we did not consider any situations which cause difficulties in calculation of a homography matrix and an essential matrix. Thus future work includes mathematical approach for stable image registration in various situations and also experiments with actual aerial images.

ACKNOWLEDGMENT

The present study was supported in part by a Grant-in-Aid for Scientific Research (A) (26249029).

TABLE III
GRIC SCORE AT FRAMES 1, 5, 9, AND 13. PLEASE NOTE THAT SMALLER GRIC SCORES MEAN BETTER EVALUATION VALUES.

Frame ID	1	5	9	13
Homography-based method	**80540.4**	**78958.0**	84097.0	66096.4
Essential-based method	82475.7	79286.8	**81813.2**	**65229.5**

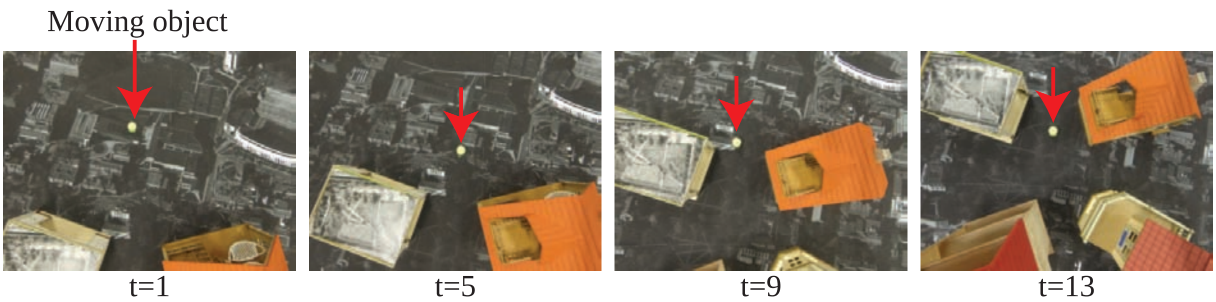


Fig. 3. Examples of captured images.

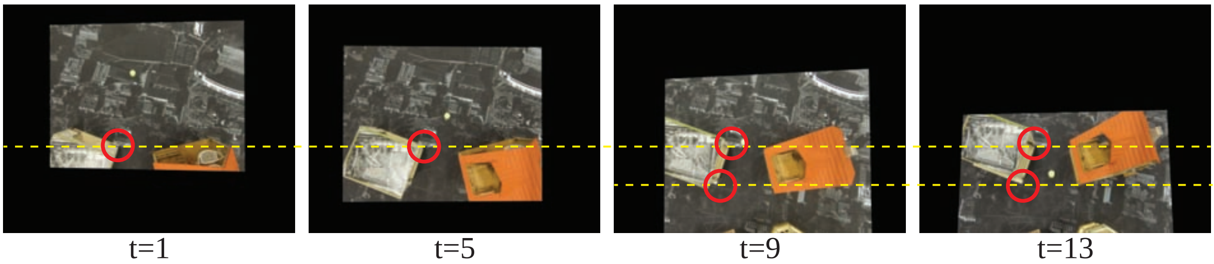


Fig. 4. Example images after image stabilization by the proposed method.

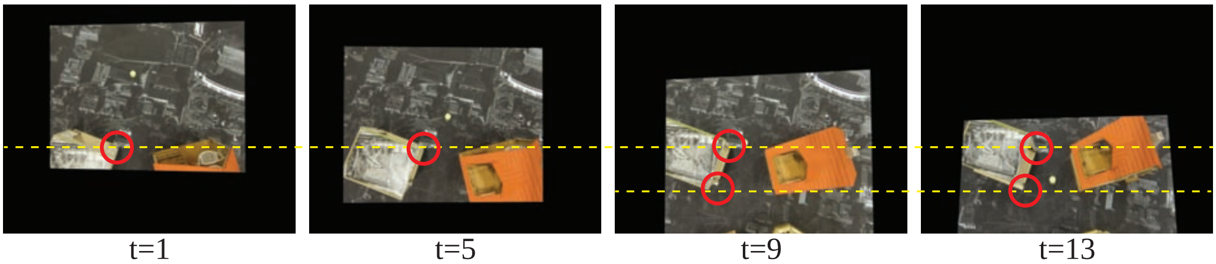


Fig. 5. Example images after image stabilization by the homography-based method

REFERENCES

[1] S. Ali and M. Shah. Cocoa - tracking in aerial imagery. *SPIE Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications*, 2006.

[2] S. Bhattacharya, H. Idrees, I. Saleemi, S. Ali, and M. Shah. Moving object detection and tracking in infra-red aerial imagery. *Machine Vision Beyond Visible Spectrum, Springer*, 2010.

[3] T. Botterill, M. Mills, and R. Green. Fast ransac hypothesis generation for essential matrix estimation. *Int. Conf. on Digital Image Computing Techniques and Applications (DICTA)*, pages 561–566, 2011.

[4] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74:59–73, 2007.

[5] O. Chum. Two-view geometry estimation by random sample and consensus. *Ph.D. dissertation, Czech Technical University*, 2005.

[6] R. Hartley and A. Zisserman. Multiple view geometry in computer vision, 2nd ed. *Cambridge University Press*, 2004.

[7] Y. Iwashita and R. Kurazume. Stable image registration for people tracking from the sky. *Emerging Security Technologies*, 2015.

[8] Y. Iwashita, M. Ryoo, T. Fuchs, and C. Padgett. Recognizing humans in motion: Trajectory-based aerial video analysis. *British Machine Vision Conference (BMVC)*, 2013.

[9] C. Lin, S. Pankanti, K. Ramamurthy, and A. Aravkin. Adaptive as-natural-as-possible image stitching. *Int. Conf. on Computer Vision and Pattern Recognition*, pages 1155–1163, 2015.

[10] M. Molinier, T. Hame, and H. Ahola. 3d-connected components analysis for traffic monitoring in image sequences acquired from a helicopter. *Lecture Notes in Computer Science*, 2005.

[11] D. Nister. An efficient solution to the five-point relative pose problem. *TPAMI*, 26:756–770, 2004.

[12] C. Olson, A. Ansar, and C. Padgett. Robust registration of aerial image sequences. *Int. Sym. on Visual Computing*, pages 325–334, 2009.

[13] Z. Shragai, S. Barnea, S. Filin, G. Zalmanson, and Y. Doytsher. Automatic image sequence registration based on a linear solution and scale invariant keypoint matching. *International Archives of Photogrammetry and Remote Sensing*, pages 5–1, 2005.

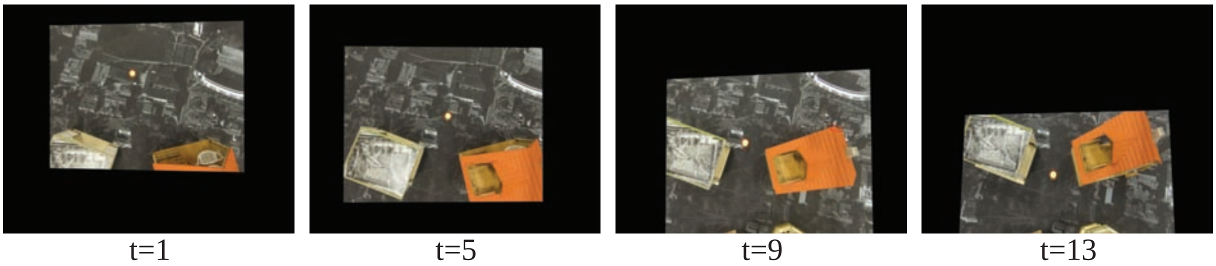


Fig. 7. Example results of object detection by the proposed method. Detected objects are marked with red lines.

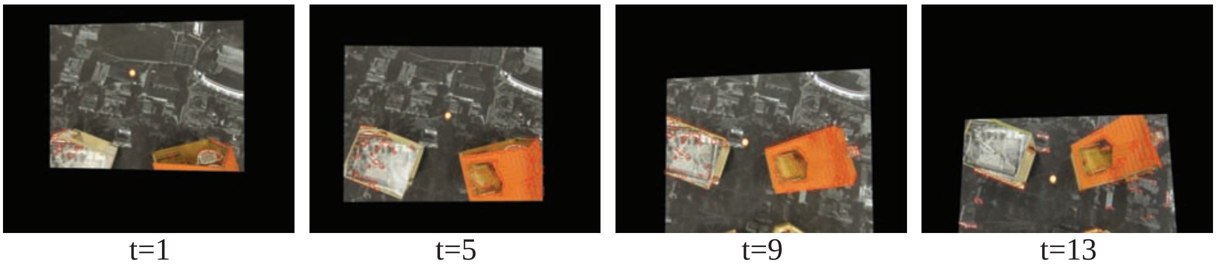


Fig. 8. Example results of object detection by the homography-based method. Detected objects are marked with red lines.

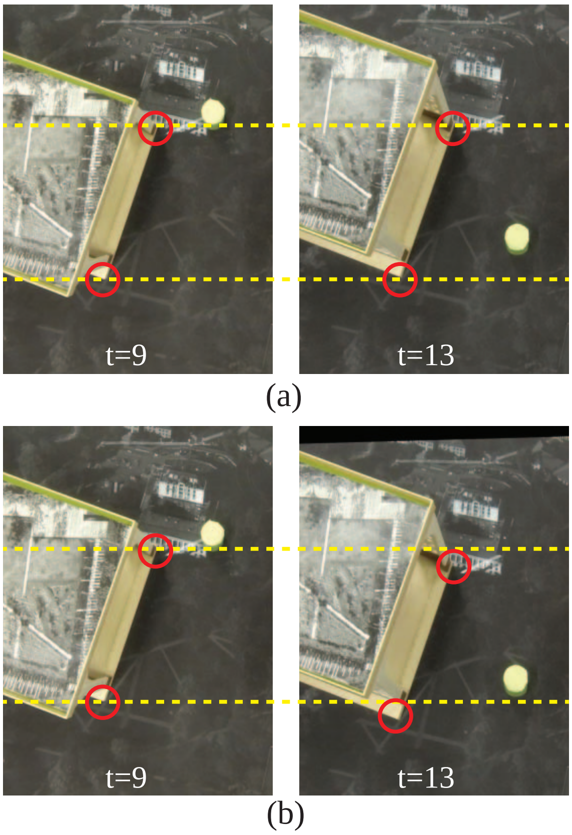


Fig. 6. Comparison between (a) stabilization results by the proposed method and (b) those by the homography-based method.