Vision-based Absolute Localization for Unmanned Aerial Vehicles

Aurélien Yol, Bertrand Delabarre, Amaury Dame, Jean-Émile Dartois and Eric Marchand

***Abstract*— This paper presents a method for localizing an Unmanned Aerial Vehicle (UAV) using georeferenced aerial images. Easily maneuverable and more and more affordable, UAVs have become a real center of interest. In the last few years, their utilization has significantly increased. Today, they are used for multiple tasks such as navigation, transportation or vigilance. Nevertheless, the success of these tasks could not be possible without a highly accurate localization which can, unfortunately be often laborious. Here we provide a multiple usage localization algorithm based on vision only. However, a major drawback with vision-based algorithms is the lack of robustness. Most of the approaches are sensitive to scene variations (like season or environment changes) due to the fact that they use the Sum of Squared Differences (SSD). To prevent that, we choose to use the Mutual Information (MI) which is very robust toward local and global scene variations. However, dense approaches are often related to drift disadvantages. Here, we solve this problem by using georeferenced images. The localization algorithm has been implemented and experimental results are presented demonstrating the localization of a hexarotor UAV fitted with a downward looking camera during real flight tests.**

I. Introduction

Initially introduced by the US Army during World War I (1917), UAVs were, for a long time, considered both very unreliable and inaccurate in term of maneuverability and navigation capabilities. Mostly used for military tasks claimed to be too risky or dangerous for humans, they were referred as the "Eyes of the Army" [19]. Today, they have reached such a level of sophistication that their usage and application have become a real technological challenge. Wide and expanding, UAVs applications are commonly associated with exploration, inspection or surveillance tasks. A deciding factor for the success of these tasks is the localization or navigation capability of the drone. Unfortunately, this can be particularly difficult when dealing with low-quality Inertial Measurement Unit (IMU), Global Positioning System (GPS), or when flying in bad weather conditions (which decreases the quality of the GPS navigation). Even though some approaches for robust Inertial/GPS localization have been studied [18], lack of precision and signal interferences still have to be considered. An alternative to avoid these problems is to mount a vision system on the UAV which can provide a temporary alternative to IMU/GPS navigation unit.

Images are offering a wealth of information, and video cameras, if they are not already present, are cheap and light enough to be mounted on board. While vision-based localization is well understood, it is often simultaneously associated with mapping (SLAM). However, with respect to vision-based SLAM for grounded-robot applications that has been largely explored, using such approaches for aerial vehicles and outdoor applications is still topical [7], [3], [14], [1], [5]. Furthermore, pure vision-based SLAM does not provide an absolute localization, may be subject to drift, and is prone to errors due to wrong estimation of the scale factor. Obviously, such issues can be alleviated when fusion with IMU or GPS is achieved.

Fig. 1. Hexacopter used for the experiments

In contrast to SLAM approaches (that use both localization and mapping), our method is a direct pose estimation which bypasses the mapping process. It can be considered as an image registration problem between the acquired image and a mosaic of georeferenced images (which can be provided by any organism such as US Geological Survey (USGS) or Institut national de l'information géographique et forestière (IGN)). Image registration is a fundamental step in vision-based autonomous UAVs applications such as localization, navigation [8] and safe landing [22], [2]. In this paper, our goal is to localize the UAV from a set of georeferenced images. This approach has already been studied in [24] which proposes a solution for a manned aircraft localization by matching georeferenced images with a sequence taken from an on-board downward looking camera. Localization of the UAV can also be done by using a mosaicing process, as in [6] where they estimate the position of the vehicle from motions. However, most of the image registration techniques rely on visual features such as keypoints [13], edges [8], or MSER [17].

Unfortunately, vision-based approaches depending on visual features are often very sensitive to illumination changes and environmental modifications. To provide accurate position information, the image processing technique used

A. Yol is with Inria Rennes, Lagadic team, Rennes, France.
B. Delabarre and E. Marchand are with Université de Rennes 1, IRISA, INRIA Lagadic team, Rennes, France. Eric.Marchand@irisa.fr
A. Dame is with University of Oxford, Oxford, United-Kingdom.
J.-E. Dartois is with Inria Rennes, DiverSE team, Rennes France.

remains crucial. Here, we use a dense or direct method. Also called template based registration, it has barely been explored when dealing with UAV localization [16]. The objective of this method is to optimize, for the camera motion, a similarity function between the image acquired by the camera and a reference template (here one of the georeferenced images). The choice of the similarity function is fundamental especially considering outdoor scenes. First, as in [16], is the Sum of Squared Differences (SSD) [4], which compares the luminance of each pixels. Unfortunately, this method is very sensitive to scene variations such as season, weather or environment changes. More advanced techniques use the Sum of Conditional Variance (SCV) [21] or the Normalized Cross Correlation (NCC) [23]. They are more robust than the SSD but still sensitive to local intensity variations. In this paper, we use the Mutual Information (MI) [25], [9]. Contrary to the previous optimization functions, MI, even if more complex to compute, is robust to local and global changes since it is based on the amount of information shared by the current and reference (georeferenced) images.

The paper is organized as follows. First, we introduce the absolute localization problem applied for UAVs navigation. Then the main principles of differential template tracking are recalled and the warping functions used for the localization are introduced. Finally, the experimental results show the robustness of our approach in difficult environment conditions both on specific images and real flight sequences (for which a comparison with GPS logs is available).

II. Absolute localization: overview

As already stated our goal is to perform 6 degrees of freedom (dof) localization (or navigation) of the UAV. Classical vision-based approaches to absolute localization rely either on a pose estimation scheme that requires a 2D-3D registration process, or on the integration of the camera motion that induces drift in the navigation process. Here we also rely on the integration of the camera motion but the possible drift effects are mitigated by considering that a set of georeferenced images is available. Also called georeferenced mosaic, it is built in an off-line processing using orthoimages that are aerial photographs or satellite images geometrically corrected and perfectly geolocalized to be usable as a map. In our case, it is represented by a mosaic of satellite images (see right of Figure 2) where a frame $\mathcal{R}_{W_{i,j}}$ is attached to each image of the mosaic. The localization of $\mathcal{R}_{W_{i,j}}$ is known with respect to a global (or world) reference frame $\mathcal{R}_W$ and each georeferenced image $I^*_{W_{i,j}}$ (template) is associated to this frame. Information like each pixel size in meters, orientation and localization of the template with relation to the world frame allows us to georeference each $\mathcal{R}_{W_{i,j}}$ with respect to $\mathcal{R}_W$.

The estimation of the absolute UAV localization consists in finding the position of the camera in the world frame ${}^c\mathbf{M}_W$. It is given by:

$$ {}^c\mathbf{M}_W = {}^c\mathbf{M}_{W_{i,j}} {}^{W_{i,j}}\mathbf{M}_W \quad (1) $$

where ${}^{W_{i,j}}\mathbf{M}_W$ is the relative position, with respect to the world frame, of the frame associated to each orthoimage. It is known by construction of the orthomosaic and is given by the provider of the orthomosaic (government agencies such as USGS or IGN, or commercial companies such as SpotImage). ${}^c\mathbf{M}_{W_{i,j}}$ is the estimated position of the camera with respect to the mosaic patch the closest to the current position. To compute ${}^c\mathbf{M}_{W_{i,j}}$, we propose to use a differential template registration method (see section III). In the experiments, the UAV flies at an altitude of approximately 150m. It allows us to consider that the observed scene is almost planar (which, obviously is a rough approximation due to the presence of buildings). In consequence, an adequate representation of the UAV motion is the homography ${}^c\mathbf{H}_{W_{i,j}}$, which can be subsequently decomposed in order to recover the UAV 3D relative position with respect to the frame associated with the closest orthoimage (see details in section III-D).

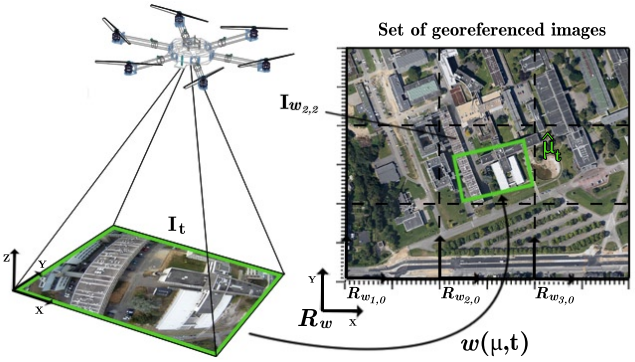


Fig. 2. Absolute localization of the UAV using a set of georeferenced images.

III. Differential template-based image registration

As stated, our absolute localization problem relies on the estimation of a 2D transformation between a georeferenced image and the current image acquired by the camera. From this transformation we extract the relative position (translation and rotation) between the current viewpoint and the position of the camera that acquired the reference image.

The homography can be estimated through point correspondences [11], [20]. However, in our case and considering that the camera motion between two frames is small we choose to rely on a differential tracker. Differential image alignment [15], [4] (or template tracking) is a class of approaches based on the optimization of an image registration function.

A. Differential tracking

The goal is to estimate the transformation $\boldsymbol{\mu}$ between a georeferenced image I^* and the current image. Note that in our absolute navigation context, I^* is the patch $I^*_{W_{i,j}}$ of the georeferenced mosaic that is the closest from the current

estimated UAV position. In the case of a similarity function f, the problem can be written[1] as :

$$\widehat{\boldsymbol{\mu}}_t \quad = \quad \arg\max_{\boldsymbol{\mu}} f(I^*, w(I_t, \boldsymbol{\mu})) \tag{2}$$

where we search the transformation $\widehat{\boldsymbol{\mu}}_t$ that maximizes the similarity between the template I^* and the warped current image I_t. In the case of a dissimilarity function the problem would be simply inverted in the sense that we would search the minimum of the function f.

To solve the maximization problem, the assumption made in the differential image registration approaches is that the displacement of the object between two consecutive frames is quite small. The previous estimated transformation $\widehat{\boldsymbol{\mu}}_{t-1}$ can therefore be used as the first estimation of the current one to perform the optimization of f and incrementally reach the optimum $\widehat{\boldsymbol{\mu}}_t$.

B. Similarity measure

One essential choice remains the one of the alignment function f. One natural solution is to choose the function f as the sum of squared differences of the pixel intensities between the reference image and the transformed current image [4]:

$$\widehat{\boldsymbol{\mu}}_t \quad = \quad \arg\min_{\boldsymbol{\mu}} \sum_{\mathbf{x} \in I^*} \left(I^*(\mathbf{x}) - I_t(w(\mathbf{x}, \boldsymbol{\mu}))\right)^2 \tag{3}$$

where the summation is computed on each point $\mathbf{x}$ of the reference image. As suggested by its definition, this dissimilarity function is very sensitive to illumination variations. In our case, variation between the current and reference images may go far beyond a simple illumination variation. When acquired at different dates or seasons, reference and current images may be very different and the SSD may prove to be an inefficient dissimilarity function.

NCC and ZNCC have shown some very good results in alignment problems [12] but in this paper we propose to define our alignment function as the mutual information [25], [9]. Originating from the information theory, MI is a measure of statistical dependency between two signals (or two images in our case) that is robust to large variations of appearance. In that case, the problem can be formulated as:

$$\widehat{\boldsymbol{\mu}} \quad = \quad \arg\max_{\boldsymbol{\mu}} \mathrm{MI}\left(I^*(\mathbf{x}), I(w(\mathbf{x}, \boldsymbol{\mu}))\right). \tag{4}$$

Rather than comparing intensities, as considered by the SSD, MI is the quantity of information shared between two random variables. The mutual information of two images I and I^* is then given by the following equation:

$$\mathrm{MI}(I, I^*) = h(I) + h(I^*) - h(I, I^*) \tag{5}$$

where the entropy $h(I)$ is a measure of variability of a random variable, here the image I and $h(I, I^*)$ is the joint entropy of two random variables which can be defined as the variability of the couple of variables (I, I^*).

[1]For the purpose of clarity, the warping function w is here used in an abuse of notation to define the overall transformation of the image I by the parameters $\boldsymbol{\mu}$. Its proper formulation should be preferred using $w(\mathbf{x}, \boldsymbol{\mu})$ to denote the position of the point $\mathbf{x}$ transformed using the parameter $\boldsymbol{\mu}$.

This can be simply computed by image histogram manipulation. If r are the possible values of I and $p_I(r) = P(I = r)$ is the probability distribution function of r, then the Shannon entropy $h(I)$ of I is given by:

$$h(I) = -\sum_r p_I(r) \log\left(p_I(r)\right). \tag{6}$$

The probability distribution function of the gray-level values is then simply given by the normalized histogram of the image I. The entropy can therefore be considered as a dispersion measure of the image histogram. Following the same principle, the joint entropy $h(I, I^*)$ of two random variables I and I^* can be computed as:

$$h(I, I^*) = -\sum_{r,t} p_{II^*}(r, t) \log\left(p_{II^*}(r, t)\right) \tag{7}$$

where r and t are respectively the possible values of the variables I and I^*, and $p_{II^*}(r, t) = P(I = r \cap I^* = t)$ is the joint probability distribution function. In our problem I and I^* are images. Then r and t are the gray-level values of the two images and the joint probability distribution function is a normalized bi-dimensional histogram of the two images. As for the entropy, the joint entropy corresponds to a dispersion measure of the joint histogram of (I, I^*).

If this expression is combined with the previously defined differential motion estimation problem, we can consider that the image I is depending on the displacement parameters $\boldsymbol{\mu}$. If we use the same warp function notation as in section III-A, MI can thus be written with respect to $\boldsymbol{\mu}$:

$$\mathrm{MI}(w(I, \boldsymbol{\mu}), I^*) = h(w(I, \boldsymbol{\mu})) + h(I^*) - h(w(I, \boldsymbol{\mu}), I^*). \tag{8}$$

C. Optimization of the similarity measure

Multiple solutions exist to compute the update of the current displacement parameters and perform the optimization. Baker and Matthews showed that two formulations were equivalent [4] depending on whether the update is acting on the current image or the reference. The former is the direct compositional formulation which considers that the update is applied to the current image. A second equivalent formulation, that is considered in our problem, is the inverse compositional formulation which considers that the update modifies the reference image, so that, at each iteration k, $\Delta\boldsymbol{\mu}$ is chosen to maximize:

$$\Delta\boldsymbol{\mu}^k = \arg\max_{\Delta_{\boldsymbol{\mu}}} f(w(I^*, \Delta\boldsymbol{\mu}), w(I_t, \boldsymbol{\mu}^k)). \tag{9}$$

In this case the current parameters is updated using:

$$w(\ w^{-1}(\mathbf{x}, \Delta\boldsymbol{\mu}^k), \boldsymbol{\mu}^k) \rightarrow w(\mathbf{x}, \boldsymbol{\mu}^{k+1}). \tag{10}$$

Considering the mutual information as the similarity function, we have:

$$\Delta\boldsymbol{\mu}^k = \arg\max_{\Delta_{\boldsymbol{\mu}}} \mathrm{MI}\left(I^*(w(\mathbf{x}, \Delta\boldsymbol{\mu})), I(w(\mathbf{x}, \boldsymbol{\mu}^k))\right). \tag{11}$$

In the inverse compositional formulation [4], since the update parameters are applied to the reference image, the

derivatives with respect to the displacement parameters are computed using the gradient of the reference image. Thus, these derivatives can be partially precomputed and the algorithm is far more efficient.

Estimating the update using a first-order optimization method such as a steepest gradient descent is not adapted. Such non-linear optimizations are usually performed using a Newton's method that assumes the shape of the function to be parabolic. Newton's method uses a second order Taylor expansion at the current position $\boldsymbol{\mu}^{k-1}$ to estimate the update $\Delta\boldsymbol{\mu}$ required to reach the optimum of the function (where the gradient of the function is null). The same estimation and update are performed until the parameter $\boldsymbol{\mu}^k$ effectively reaches the optimum. The update is estimated following the equation:

$$\Delta\boldsymbol{\mu} = -\mathbf{H}_{MI}^{-1}\mathbf{G}_{MI}^{\top} \tag{12}$$

where $\mathbf{G}_{MI}$ and $\mathbf{H}_{MI}$ are respectively the gradient and Hessian matrices of the mutual information with respect to the update $\Delta\boldsymbol{\mu}$. Following the inverse compositional formulation defined in equation (9) those matrices are equal to:

$$\mathbf{G}_{MI} = \frac{\partial \text{MI}(w(I^*, \Delta\boldsymbol{\mu}), w(I, \boldsymbol{\mu}))}{\partial \Delta\boldsymbol{\mu}} \tag{13}$$

$$\mathbf{H}_{MI} = \frac{\partial^2 \text{MI}(w(I^*, \Delta\boldsymbol{\mu}), w(I, \boldsymbol{\mu}))}{\partial \Delta\boldsymbol{\mu}^2}. \tag{14}$$

The details of the calculation of equations (13) and (14) can be found in [10] and an optimized version that allows a second order optimization of the mutual information as been recently proposed in [9].

D. Warp functions

Along with the choice of the similarity function, the choice of an adequate warping function is fundamental considering that the UAV underwent an arbitrary 3D motion. It has to be noted that there does not exist a 2D motion model that accounts for any kind of 3D camera motion.

With respect to the camera altitude, ground altitude variations are often small. In that case, the homography, that is able to account for any 3D motion when the scene is planar, is a good candidate for the transformation.

A point $\mathbf{x}_t$, expressed in homogeneous coordinates $\mathbf{x}_t = (x_t, y_t, 1)$, is transfered in the other image using the following relation:

$$w(\mathbf{x}_t, \boldsymbol{\mu}) = \mathbf{H}\mathbf{x}_t \tag{15}$$

where $\mathbf{H}$ is the homography. This homography can be linked to the 3D camera motion $(\mathbf{t}, \mathbf{R})$ by:

$$\mathbf{H} = (\mathbf{R} + \frac{\mathbf{t}}{d}\mathbf{n}^T) \tag{16}$$

where $\mathbf{n}$ and d are the normal and the distance to the reference plane (here the ground) expressed in the camera frame. Knowing the homography $\mathbf{H}$ it is possible to decompose it in order to retrieve the absolute displacement of the camera [11].

Alternatively, if the camera is a real nadir camera (ie, the image plane and "flat" ground are almost parallel), the homography appears to be over parameterized. The s$\mathbf{Rt}$ model accounts for 3D motions when the image and scene planes remain parallel (ie, optical axis perpendicular to the ground). It is a combination of a 2D rotation $\mathbf{R}_{2d}$ (which accounts for the rotation around the optical axis), a translation $\mathbf{t}_{2d}$ and a scale factor $\mathbf{s}$. The warp of a point $\mathbf{x}_t = (x_t, y_t)^T$ is then given by:

$$w(\mathbf{x}_t, \boldsymbol{\mu}) = s\mathbf{R}_{2d}\mathbf{x}_t + \mathbf{t}_{2d}. \tag{17}$$

As for the homography, from this simplified form, we can also retrieve the absolute 3D displacement of the camera. In the experiments the camera is mounted on the UAV with a IMU piloted camera which allows the camera to be in a Nadir configuration where pitch and roll motions are counterbalanced. We therefore rely on this $s\mathbf{Rt}$ model.

IV. Experimental results

The proposed method has been validated using flight-test data acquired with a GoPro camera mounted on an hexarotor UAV (Figure 1). The UAV size is approximately 70cm. It is powered by six 880kv motors. It also has an embedded GPS which is only used as "ground-truth" to assess the computed localization. Our localization process is currently done off board via an Intel Xeon 2.8GHz.

The flight data were collected above the Université de Rennes 1 campus. The resolution of the georeferenced mosaic used for image registration is approximately 0.2 meter per pixel. All the experiments used a motorized Nadir system where pitch and roll motions of the UAV are counterbalanced via a brushless gimbal. Flights altitude is approximately 150 meters high ($\pm$25m). Perspective effects are limited and scene is assumed to be planar. These particularities make the usage of the $s\mathbf{Rt}$ warping function more suitable than the homography which is overparametrized. As a result, 4 degrees of freedom localization are estimated and defined by the geometric coordinates of the UAV, its altitude and heading.

A. Analysis of the similarity measure

In order to assess the choice of MI as our similarity function, a simple experiment was realized that compare the SSD and MI in extreme conditions.

We consider a reference template of 650$\times$500 pixels extracted from Google Earth in the region of Edmonton, CA in summer and a current image of the same scene in winter. A translation of $\pm$20 pixels along each axis is considered and the values of the SSD and MI are computed between the reference and current patch. The shapes of the cost functions (equation(3) and (4)), along with the considered images, are shown on Figure 3. The result is consistent with our choice. MI shows a well defined maximum at the expected position proving that it is perfectly suited for localization, contrary to the SSD which is tricked by the differences resulting of seasonal changes, lighting conditions or environmental modifications.

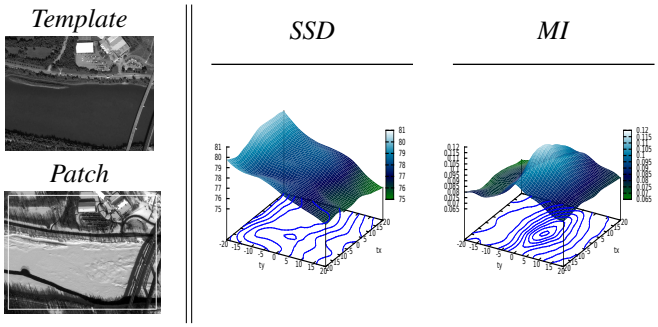


Fig. 3. Comparison of the similarity functions. The box drawn on the current patch shows the position resulting of $tx = 0$ and $ty = 0$ pixel.

B. Validation of motion estimation

In this experiment we try to validate the motion obtained from the MI-based registration approach. As explained in section II, we want to determine the transformation between a reference image (Figure 4-a) and the current view of the camera (Figure 4-b). Let us note that here, the patch of the georeferenced mosaic (obtained here from satellite images), $I^*_{W_{i,j}}$ in section II, is represented by the green rectangle on Figure 4-a. The resulting position estimation, that has been computed thanks to the registration step using a s$\mathbf{Rt}$ displacement model, is visible on Figure 4-c. The fact that the template used on (a) is perfectly retrieved on (c) secures our ambitions to use the differential template-based registration in a vision-based algorithm to retrieve the localization of the UAV during an actual flight.

Once again, this experiment proves that our algorithm is robust enough to estimate the motion of the UAV by simply comparing images gathered thanks to the downward looking camera with satellite images.

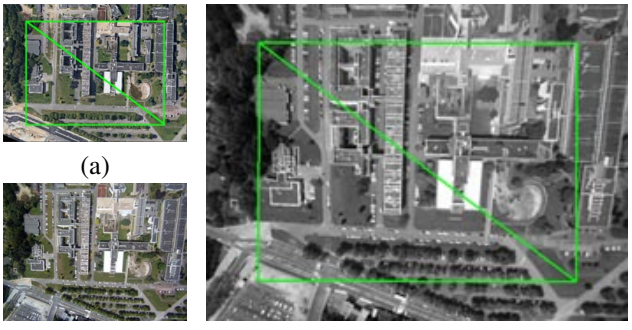


Fig. 4. Result of the motion estimated from the template-based image registration. On (a) is the template used (extracted from Google Earth), (b) is the camera view and (c) the result of the registration.

C. Real flight experiments

In order to test our algorithm on real sequences, we proceeded to fly the presented UAV following the 695m path shown in green on Figure 5 and recorded flight images and localization data. The geographic localization and the altitude have been logged through a GPS. The flight lasted approximately 4 minutes, and the UAV average speed was about 3 m/s. The localization initialization in the very first image is done using GPS data (as stated vision-based navigation is used to provide temporary alternative to IMU/GPS navigation unit). Neither take-off nor landing phase is considered here.

Fig. 5. Ground-Truth path obtained through the GPS and geo-referenced mosaic with simulated fog used for the experiment.

Obtaining real flight sequences in bad weather conditions has not been conceivable (for safety issues). Consequently, to prove the robustness of our approach in such situations we have decided, in this experiment, to consider two cases. Firstly, the absolute localization has been performed with a classical georeferenced mosaic extracted from Google Earth. Then, we used the same mosaic where we have simulated a thick fog (see Figure 5) to reduce the visibility of the scene on the reference templates.

Geographic positions, originally characterized by longitude and latitude angles, have been converted in metric coordinates with a Lambert93 projection model, which is the official projection for metropolitan French maps. In the region of Rennes, France, where our data have been collected, the linear alteration induced by this projection is approximately 550mm/km. In order to compare properly the GPS altitude and the altitude computed by our algorithm, we have considered the current mosaic as 40 meters above the sea level, which is an approximate value for this area of Rennes. For readability purposes, all the data have been normalized to consider the first GPS position as reference position (see Figure 6).

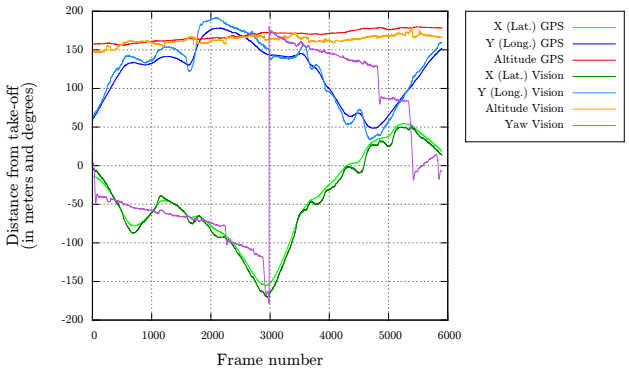


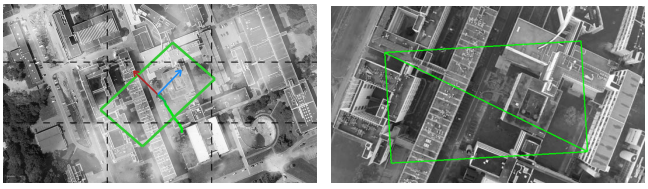
Fig. 6. Absolute localization of the UAV when using geo-referenced mosaic with fog, and GPS ground-truth.

Ground-truth localization data have been collected at 5Hz which explains the smoothness of the GPS curves on Figure 6, which is not present on the localization obtained from our vision-based approach using images acquired at 30Hz. The s$\mathbf{Rt}$ motion has been considered in this experiment. Despite the fact that perspective effects are still visible at this altitude and thus impact the navigation process, the estimated trajectory can be favorably compared to the one given by the GPS (see Figure 6). According to the computed Root-Mean-Square Deviation (RMSE) (see Figure 7), MI is slightly more accurate than the SSD on the sequence using the classical mosaic. This can be explained by the fact that the SSD is more affected by the perspective and lighting differences present on the reference. On the other case, when considering fog, the MI accuracy is not affected contrary to the SSD which failed and is not able to localize the UAV.

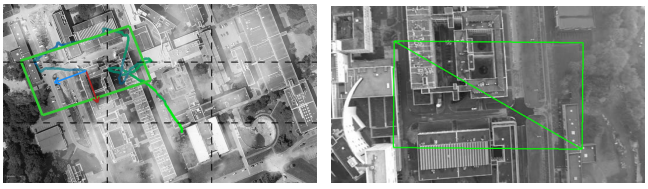
RMSE	X (Lat.)	Y (Long.)	Altitude
MI (With Fog)	6.56926m	8.2547m	7.30694m
MI	6.5653m	8.01867m	7.44319m
SSD	6.79299m	8.03995m	6.68152m

Fig. 7. Root-mean-square deviation on the different degrees of freedom between our approach and the GPS values.

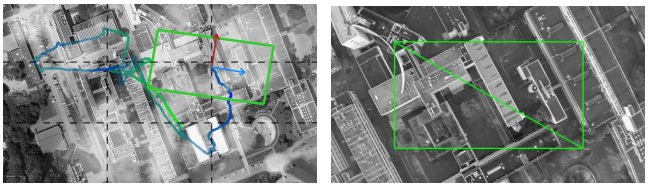
As can be seen on the Figure 6 and 8, the computed trajectory is very close from the ground-truth (see Figure 7). On the left of Figure 8, we can see the current estimated trajectory, the current heading and the reference patch. On the right, Figure 8 shows the registered patch on the current view. Finally let us note that neither filtering process, such as a Kalman filter, nor fusion with other sensors have been considered here. Results from Figure 6 are rough data obtained from a vision only method.



Frame 500



Frame 3500



Frame 5900

Fig. 8. Result of the localization, plus current template of reference (left). Result of the image registration (right).

V. Conclusions

We introduced a new way to localize a UAV using a vision process only. Our approach uses a robust image registration method based on the mutual information. By using a geo-referenced mosaic, drift effects, a major problem with dense approaches, can be avoided. Plus, it gives us the possibility to estimate the absolute position of the vehicle from its estimated 2D motions. A next phase would be to integrate our approach in a global estimation framework (while including IMU data) and to realize a real-time on board localization when sufficient CPU power will be available on the UAV.

References

[1] S. Ahrens, D. Levine, G. Andrews, J.P. How. Vision-based guidance and control of a hovering vehicle in unknown, GPS-denied environments. In *ICRA'09*, pp. 2643-2648, Kobe, 2009.

[2] G. Anitha, R.N. Gireesh Kumar. Vision based autonomous landing of an unmanned aerial vehicle. *Procedia Engineering*, vol 38, 2012.

[3] J. Artieda, et al. Visual 3D SLAM from UAVs. *J. of Intelligent and Robotic Systems*, 55(4-5):299-321, 2009.

[4] S. Baker, I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221-255, 2004.

[5] M. Blösch, S. Weiss, D. Scaramuzza, R. Siegwart. Vision based mav navigation in unknown and unstructured environments. *IEEE ICRA'10*, pp. 21-28, Anchorage, AK, May 2010.

[6] F. Caballero, L. Merino, J. Ferruz, A. Ollero. Improving vision-based planar motion estimation for unmanned aerial vehicles through online mosaicing. *ICRA 2006*, pp. 2860-2865, May 2006.

[7] F. Caballero, L. Merino, J. Ferruz, A. Ollero. Vision-based odometry and SLAM for medium and high altitude flying UAVs. *Unmanned Aircraft Systems*, pp. 137-161. Springer, 2009.

[8] G. Conte P. Doherty. An integrated uav navigation system based on aerial image matching. *IEEE Aerospace Conf.*, Big Sky, March 2008.

[9] A. Dame, E. Marchand. Second order optimization of mutual information for real-time image registration. *IEEE T. on Image Processing*, 21(9):4190-4203, September 2012.

[10] N.D.H. Dowson, R. Bowden. A unifying framework for mutual information methods for use in non-linear optimisation. *ECCV'06*, pp. 365-378, June 2006.

[11] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2001.

[12] M. Irani, P. Anandan. Robust multi-sensor image alignment. In *ICCV'98*, pp. 959-966, Bombay, India, 1998.

[13] F. Kendoul, F. Fantoni, K. Nonami. Optic flow-based vision system for autonomous 3d localization and control of small aerial vehicles. *Robotics and Autonomous Systems*, 57(6-7):591-602, 2009.

[14] J.H. Kim, S. Sukkarieh. Airborne simultaneous localisation and map building. In *IEEE ICRA '03.*, volume 1, pp. 406-411, 2003.

[15] B.D. Lucas, T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81*, pp. 674-679, 1981.

[16] C. Martinez, et al. A hierarchical tracking strategy for vision-based applications on-board uavs. *Journal of Intelligent & Robotic Systems*, 72(3-4):517-539, 2013.

[17] L. Merino, et al. Vision-based multi-UAV position estimation. *IEEE Robotics Automation Magazine*, 13(3):53-62, 2006.

[18] A. Nemra, N. Aouf. Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering. *IEEE Sensors J.*, 2010.

[19] U. S. Army UAS Center of Excellence. *U.S. Army Roadmap for Unmanned Aircraft Systems, 2010-2035*. 2010.

[20] M. Pressigout, E. Marchand. Model-free augmented reality by virtual visual servoing. In *ICPR'04*, pp. 887-891, Cambridge, 2004.

[21] R. Richa, et al. Visual tracking using the sum of conditional variance. *IROS'11*, pp. 2953-2958, San Francisco, 2011.

[22] S. Saripalli, J.F. Montgomery, G. Sukhatme. Visually guided landing of an unmanned aerial vehicle. *IEEE T-RA*, 19(3):371-380, 2003.

[23] G. Scandaroli, M. Meilland, R. Richa. Improving NCC-based direct visual tracking. *ECCV'12*, pp. 442-455, 2012.

[24] DG. Sim, et al. Integrated position estimation using aerial image sequences. *IEEE PAMI*, 24(1):1-18, 2002.

[25] P. Viola, W. Wells. Alignment by maximization of mutual information. *IJCV*, 24(2):137-154, 1997.