Image-based GNSS-denied Localization for BLOS Targeting

Mahamadou Idrissa, Xavier Neyt
Royal Military Academy
AV. de la Renaissance 30
B-1000 Brussels
BELGIUM

Mahamadou.IdrissaSeyni@rma.ac.be, Xavier.Neyt@rma.ac.be

ABSTRACT

The BLOS system for target designation depends on the accuracy with which the UAV's position and attitude are known. However, relying solely on the GNSS can be risky. The GNSS signals can be lost for a variety of reasons such as jamming, multipath reflections caused by terrain profile, weather conditions, etc. We address the framework of the GNSS denied navigation system and demonstrated the feasibility of the vision-based system involving the UAV image and a pre-existing geo-referenced image. The approach is a pipeline of algorithms for automatic image matching, ground control point extraction and platform orientation calibration. Experimentation and evaluation on two datasets showed that the solution is promising in terms of UAV positioning system.

1.0 INTRODUCTION

BLOS (Beyond Line of Sight) targeting enables firing at a target that is not in direct line of sight. A forward observer or in this case a UAV (Unmanned Aerial Vehicle) is used to get a direct line of sight and the obtain coordinates of the target. Such system for target designation depends on the accuracy with which the UAV's position and attitude are known. However, the GNSS signal is vulnerable to electromagnetic environmental effects mainly because the received power level is very low. The signal can be perturbed or even lost due to numerous situations such as multi-path reflections caused by nearby obstacles, weather conditions, interference by signals from other sources. In addition, intentional jamming or meaconing [1], [2] attack have become as a major threat for UAV navigation systems. Jamming attack works by flooding the receiver with too much power that masks completely the original signal, while the meaconing attack is the broadcasting of a GNSS signal at higher power level with inaccurate information.

A standard UAV navigation system often relies on GNSS data and an IMU (Inertial Measurement Unit). In the GNSS-denied scenarios, pure inertial navigation drifts rapidly. An alternative to address this problem is to combine the inertial module with information from computer vision. However, without any absolute position reference, the estimated position of the UAV will always suffer from a drift. One technique to solve this drifting problem could be the Simultaneous Localisation and Mapping (SLAM) [3] used in grounded-robot applications. Using this technique, the robots are able to draw a map and estimate their localisation inside that map. Loop closure detection is needed to validate the robot's estimated position in the map and correct it if necessary. However, this loop closure detection module is a limiting factor for many outdoor UAV tasks.

In this work, we discuss a framework for GNSS-denied localization and demonstrate the feasibility of an image matching-based method using the live UAV image and a pre-existing geo-referenced image. The matching allows automatic extraction of ground control points (GCPs) between the images acquired by the UAV and the reference image. These GCPs are then used to estimate the position and attitude of the UAV. However, it is difficult to match two images of the same location acquired from different sources and at

different times. To explore all possibilities in terms of resolution differences and viewpoint distortions, the UAV images are down-sampled into a pyramid of different resolutions. Different rotations and tilt angles are applied to simulate viewpoint distortions. Feature matching is then applied to all the transformed images and the reference image to establish a list of matching points. The geo-referencing information of this reference image, associated to a digital terrain model (DTM), allow exporting the matching points as GCPs. The extrinsic parameters, i.e. position and attitude of the camera are then estimated using the least squares method. The method is evaluated with two datasets. For each dataset, we calculated matching points re-projection errors according to the parameters obtained by least squares estimation.

2.0 MATCHING AND CALIBRATION

The image matching module is the most important part of this vision-based GNSS denied navigation system. It allows the automatic extraction of GCPs between the images acquired by the UAV and the pre-existing geo-referenced image. These GCPs are then used in the calibration module to estimate the position and attitude of the UAV. However, it is difficult to match two images of the same location but acquired from different sources. As reference images (satellite or orthophoto) were captured at different times, there are differences with UAV images in terms of lighting, landscape modification (e.g. cultivation field), viewing angles and of different scale. Therefore, robust and highly invariant feature matching methods are required. We then propose the use of registration methods based on SIFT (Scale-Invariant Feature Transform Lowe) [4]. This method aims to transform the content of an image into a set of feature vectors, or descriptors, which are computationally efficient with certain robustness properties to image variations (viewpoint, scale, illumination, etc.). However, large scale difference is one of the factors affecting its performance. This is the case for UAV images, which are generally of very high resolution (less than 1cm/pixel) and for satellite images typically having a resolution of 50cm.

ASIFT (Affine-SIFT) [5] is a variant of SIFT designed to match images with different viewpoints. The basic principle of this algorithm is to simulate the affine distortion caused by the camera's posture on the imaging moment through rotating and tilting the original images, and generate the affine transformation images in different conditions.

For an image pixel coordinates *(x, y)*, the affine transformation *(x', y')* is defined as follow:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \tag{1}$$

With $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ the affine transformation matrix and $\begin{bmatrix} e \\ f \end{bmatrix}$ is the translation vector.

In [5] the authors showed that this matrix A can be decomposed into several transformation matrixes using the Singular Value Decomposition (SVD):

$$A = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} 1/\cos\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \tag{2}$$

Where $\lambda > 0$ is the scale factor, $\psi \in [0, 2\pi]$ is the camera rotation angle around its optical centre while $\theta \in [0, \pi/2[$ and $\varphi \in [0, 2\pi\ ]$ are the camera viewpoint angles.

The use of this method consists in simulating all possible distortion differences (in viewpoint (θ, φ) and scale λ) between UAV image and the reference image. However, it takes a lot of time and memory to

simulate all affine images of different angles and to match multiple images at the same time. To overcome this drawback, we limited the transformations to UAV images and left the geo-referenced image unchanged. The UAV images are sub-sampled into a pyramid of 3 or 4 levels (change of λ) and each down-sampled image is tilted with 4 steps of θ ranging from 0 to π/3 and rotated with also 4 steps of angle φ. Each transformed image is then matched with the satellite image. Matching points found are merged to form the final matching results. However, several wrong matches might occur and the RANSAC (Random Sample Consensus) [6] algorithm is used to overcome this problem and refine the result. The RANSAC algorithm takes all matched points as input, formulates the homography model that incorporates the majority of these points and eliminates outliers. The matching module produces a list of tie points between the UAV image and the reference image. The geo-referencing information associated with the digital terrain model (DTM) allows exporting them as GCPs for the calibration module.

The calibration component is responsible for computing the geometric relationship between the 2D image plane (pixel positions in the image captured by the camera) and the 3D ground coordinates (world coordinates from reference image and DTM). The perspective camera model, also called the pinhole camera model is used. In this model, the relationship between a pixel p_{image} and its corresponding world coordinates P_{world} is expressed as follows:

$$p_{image} = sK[R,T]P_{world} \tag{3}$$

Where s is the scale factor; K is the camera intrinsic parameters (focal length f and camera frame centre x_0, y_0). R and T are the camera extrinsic parameters. R is a 3x3 rotation matrix derived from the UAV attitude angles ω, φ, κ while T is a vector position (X_0, Y_0, Z_0) of the UAV.

Calibration is the process to determine the tree intrinsic parameters (f, x_0, y_0) and six extrinsic parameters (ω, φ, κ, X_0, Y_0, Z_0) of a camera. We focus only on the problem of estimating the camera extrinsic parameters, and assume that its intrinsic parameters are known. Given the GCPs (from the matching module), consisting of both the world coordinates as well as the corresponding image coordinates, the determination of camera extrinsic parameters ($R_{\omega, \varphi, \kappa}$, $T_{X0, Y0, Z0}$) can be solved using Least-Square solution. Least-Square solving relies on minimizing the re-projection error between the observed camera image points and the corresponding 3D points re-projected onto the camera frame. The minimizing cost function is depicted by:

$$\arg\min_{R,T}\left(\sum_{i=1}^{n}\left\|p^{i} - p_{r}^{i}\right\|^{2}\right) \tag{4}$$

Where p^i is the camera image point with index i, whereas p_r^i is the re-projected 2D point of the corresponding 3D ground coordinates. The re-projection point p_r^i is calculated using equation (3). The integer n is the total number of GCPs.

3.0 EXPERIMENT AND EVALUATION

The framework was evaluated using two datasets. The first dataset was acquired from the NX70[1] drone equipped with a Sony DSC-RX0 camera. The sensor size is 13.21x8.89mm with a focal length of 7.70mm. The flights took place at two different test sites. One at low altitude (about 30m) with the camera fixed at an oblique view of about 30° while the other was at medium altitude (about 100m) with the camera at nadir for a vertical view. The estimated resolutions of the images are approximately 1cm and 3.6cm respectively for the two flights. The datasets are accompanied by metadata indicating the real-time recording of position and attitude of the camera. The second dataset was collected from an acquisition campaign in Cyprus with the

[1] Data provided by Novadem. Novadem is specialized in aerial robotics (UAV, Drone) intended to civilian and military markets.

UX11[2] drone. The images with a size of 5048x4228 pixels were taken at an altitude of about 120m with the camera pointed at nadir. The other datasets required by the method are the reference images available before the flights. These are mainly satellite images or orthophotos as well as DTM (Digital Terrain Model) of the areas overflown by the drones.

In the case of NX70 datasets, Figure 1 shows an example of image from onboard camera (left image) and the same area seen by a satellite image (right image). The Figure 2 shows the ability of the matching module to identify a large number of match points between the very high-resolution UAV image and the corresponding low-resolution satellite image. All images included in the datasets were correctly matched to this reference image. A few thousand match points were obtained after rejecting outliers by RANSAC filtering. The geo-referencing information of the satellite image and the DTM of the area are then used to build GCPs for the calibration module. Only a subset (50%) of these GCPs is used for calibration. The other part is used to evaluate the result.

Figure 1: On the left, an example of image taken by the UAV at low altitude and with an oblique view of 30°. On the right, the geo-referenced satellite image of the area with a rectangle showing the footprint of the UAV image.

[2] Data provided by Delair. Delair is a French drone manufacturer company.

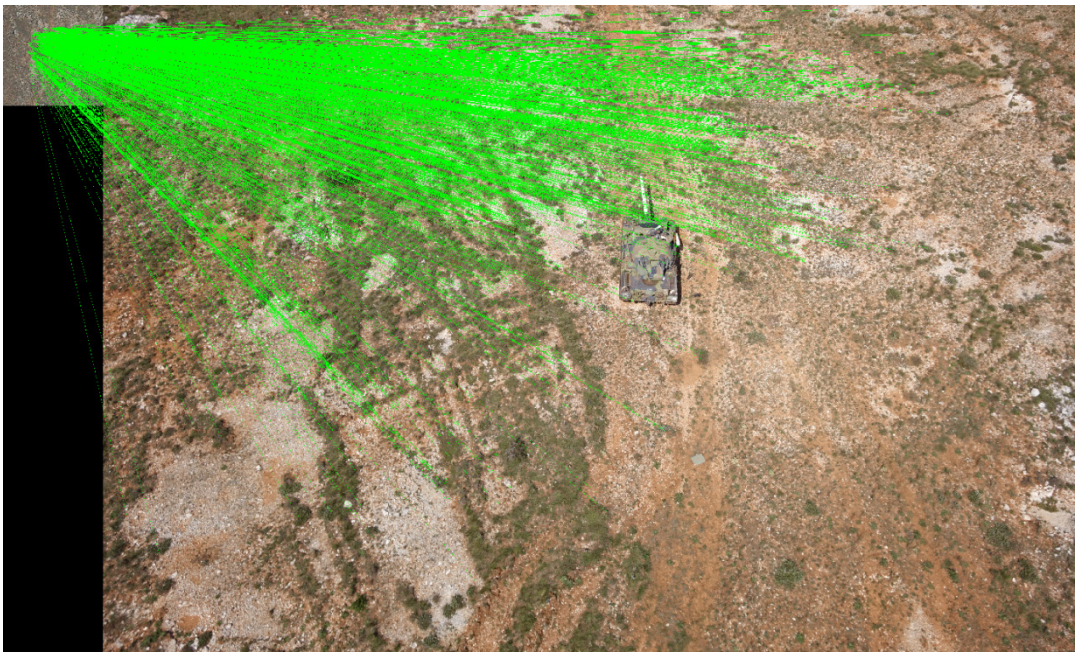

Figure 2: Example of matching using the proposed method. Green lines represent correspondences between UAV image and geo-referenced satellite image.

The evaluation consists in using the extrinsic parameters (obtained after calibration) to calculate the projections of the image points and compare them with their matched world coordinates. Table 1 shows the results of some sample images. For each sample image, the errors (mean, minimum, maximum, and median error) of projection related to the parameters obtained by calibration are computed. Images with low altitude flight with a 30° viewing angle (images D433_A353_H33 and D413_A198_H26 in the table), the mean error is globally less than 1m while the median error is in most cases below 0.5m. This median error value indicates that half of the evaluated points are accurate to within 0.5m. The same tendency is observed for the medium altitude flight (100m) with a vertical camera view (images D512_A113_H102 and D556_A198_H102). The mean error in this scenario is generally less than 2m while the median error is less than 1.5m. The Figure 3 shows their spatial distribution and it can be noted that the strong deviations appear more towards the edges of the images and particularly in the direction of the deformation due to the perspective view.

Table 1: Evaluation of positioning accuracy for low and medium altitude flights.

Ground Projection Errors (m)				
Image Ref.	**RmsErr**	**MinErr**	**MaxErr**	**MedErr**
D433_A353_H33	0.535	0.014	2.537	0.377
D413_A198_H26	0.772	0.041	3.624	0.453
D512_A113_H102	0.746	0.018	2.923	0.455
D556_A198_H102	1.680	0.025	5.313	1.262

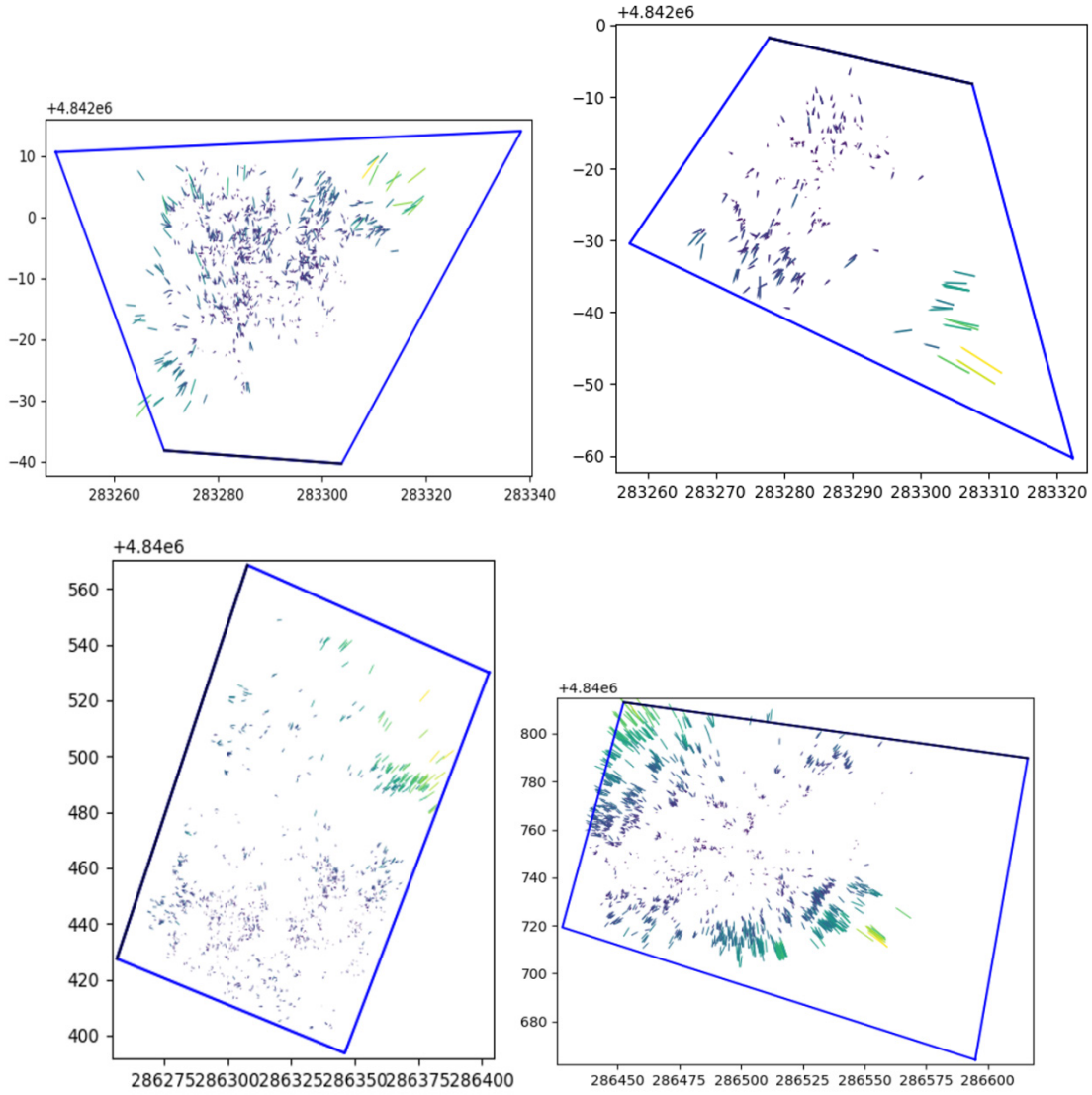

Figure 3: Graph of image footprints and spatial distribution of positioning accuracy: for low (top row) and medium (bottom row) altitude flights.

Finally, Figure 4 shows an example of image from UX11 and the orthophoto the area. The matching points obtained between the drone image and the orthophoto is depicted in Figure 5. The GCPs are derived from these points using the geo-referencing information from the orthophoto and DTM of the area. Table 2 and Table 3 present the results of the flight parameter estimation for some images. As a reminder, the flight was performed at a height of about 120m with a vertical camera view. These tables show the results obtained for the position and attitude of the camera as well as the position and attitude information from the acquisition metadata file. The results are very comparable to the information provided by the metadata. Table 4 shows the evaluations of the results obtained with the data. We can observe that for all the images tested, the accuracies are very good. The projection errors are of the order of 10cm.

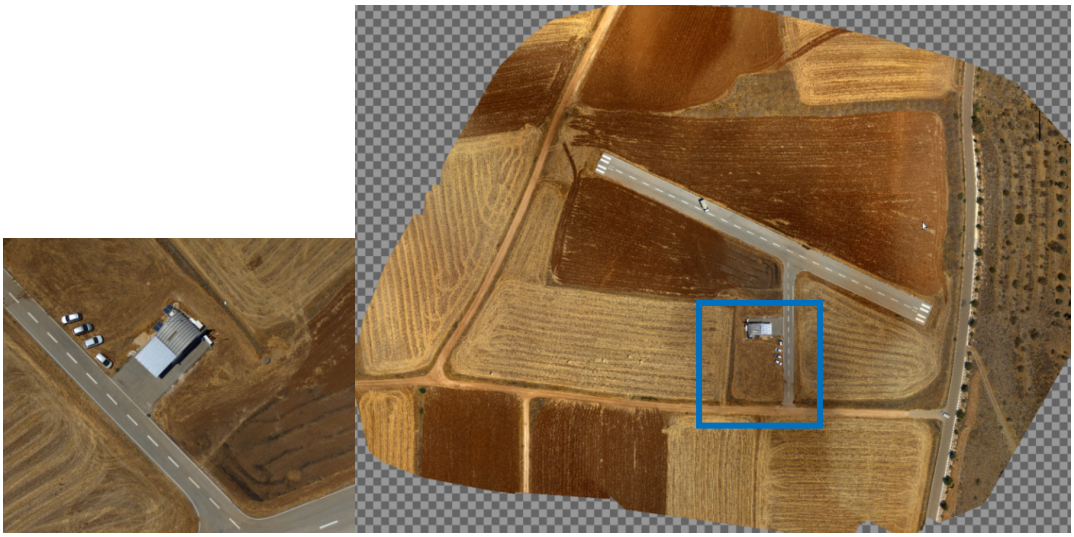

Figure 4: On the left, an example of an image taken by the Delair UAV with a vertical view of the camera. On the right, the geo-referenced orthophoto image of the area with a rectangle showing the footprint of the UAV image.

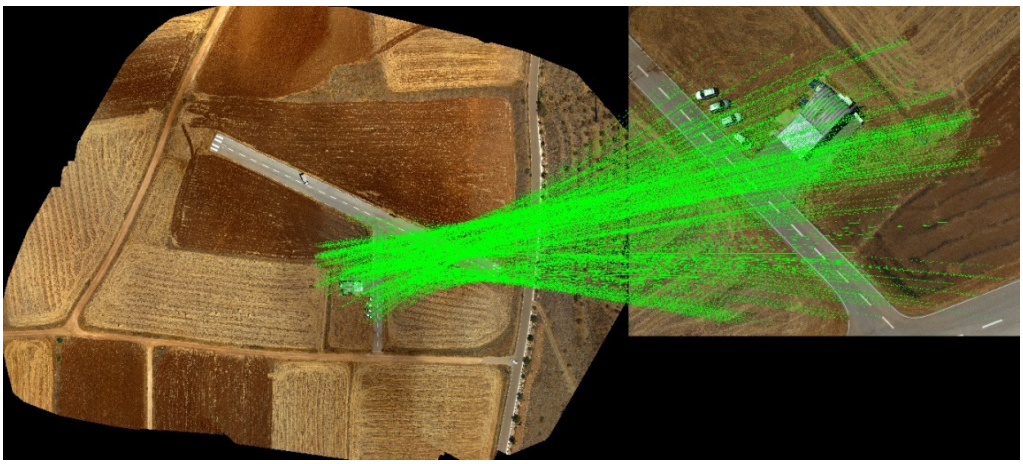

Figure 5: Example of matching using the proposed method. Green lines represent correspondences between UAV image and geo-referenced orthophoto.

Table 2: Orientation estimation results and comparison: Positions of UAV.

Position X, Y, Z (UTM - m)						
Image Ref.	Calibration			Metadata		
03_00000115	507437.549	3883550.823	433.027	507443.894	3883557.524	435.832
03_00000118	507434.224	3883523.317	432.605	507430.275	3883524.397	438.608
03_00000199	507353.936	3883525.846	435.64	507351.278	3883527.352	440.292
03_00000322	507318.326	3883665.212	434.427	507314.07	3883671.674	437.343

Table 3: Orientation estimation results and comparison: Attitudes of UAV.

Angles Roll, Pitch, Yaw (deg)						
Image Ref.	**Calibration**			**Metadata**		
03_00000115	-0.12	1.18	-135.67	-0.34	4.755	-135.229
03_00000118	2.28	5.10	-136.92	0.55	2.91	-136.867
03_00000199	-1.14	-2.11	5.98	-5.85	2.63	0.055
03_00000322	3.87	-3.31	-0.11	-6.8	-6.06	-4.66

Table 4: Evaluation of positioning accuracy.

Ground Projection Errors (m)				
Image Ref.	**RmsErr**	**MinErr**	**MaxErr**	**MedErr**
03_00000115	0.097	0.003	0.167	0.091
03_00000118	0.105	0.004	0.182	0.102
03_00000199	0.103	0.002	0.190	0.098
03_00000322	0.103	0.003	0.213	0.097

4.0 CONCLUSION

We addressed the framework of the GNSS denied navigation system and demonstrated the feasibility of a vision-based system involving the UAV image and the pre-existing geo-referenced satellite image. The approach is a pipeline of algorithms for automatic image matching, ground control points extraction and platform position and attitude estimation. Experimentation and evaluation on the NX70 and UX11 dataset combined with geo-referenced imagery and DTM (Digital Terrain Model) data showed that the solution is promising in terms of UAV positioning accuracy.

The good results obtained with the UX11 data (spatial coordinate accuracy globally in the order of 10cm) demonstrate that the solution works when favourable conditions are met:

- Drone image and orthophoto of good radiometric quality
- Drone image with very high spatial resolution
- Orthophoto with good spatial resolution
- Images taken at medium altitude (+- 100m)
- Camera oriented at nadir

ACKNOWLEDGMENTS

This work has been supported by LynkEUs project. LynkEUs is funded by European Commissions' European Defence Industrial Development Programme under the grant agreement EDIDP NGPSC-2019-014.

5.0 REFERENCES

[1] Khan, Shah Zahid, Mujahid Mohsin, and Waseem Iqbal. On GPS spoofing of aerial platforms: a review of threats, challenges, methodologies, and future research directions. PeerJ Computer Science 7 (2021): e507.

[2] Elezi E, ankaya G, Boyac A, Yarkan S. 2019. The effect of electronic jammers on GPS signals. In: 2019 16th international multi-conference on systems, signals devices (SSD). 652–656.

[3] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, An overview to visual odometry and visual SLAM: Applications to mobile robotics, Intell. Ind. Syst., vol. 1, no. 4, pp. 289–311, Dec. 2015.

[4] Lowe, David G. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60.2 (2004): 91-110.

[5] G. Yu and J. -M. Morel, A fully affine invariant image comparison method, *2009* IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 2009, pp. 1597-1600.

[6] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6):381–395, 1981.