International Conference on Robotics, Automation, Control and Embedded Systems – RACE 2015
18-20 February 2015, Hindustan University, Chennai, India.

Visual Odometry Based Absolute Target Geo-Location from Micro Aerial Vehicle

Arun Annaiyan
Automation Research Group,
Interdisciplinary Centre for Security, Reliability and Trust,
University of Luxembourg.
e-mail: Arun.annaiyan@uni.lu

Mahadeeswara Yadav
Flight Simulation and Avionics Division
Sheeba Computer Pvt. Ltd, Bangalore, India
e-mail: mahadeesyadav@yahoo.com

Miguel A.Olivares Mendez
Automation Research Group,
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg.

Holger voos
Automation Research Group,
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg.

Abstract: - **An unmanned aerial system capable of finding world coordinates of a ground target is proposed here. The main focus here was to provide effective methodology to estimate ground target world coordinates using aerial images captured by the custom made micro aerial vehicle (MAV) as a part of visual odometery process on real time. The method proposed here for finding target's ground coordinates uses a monocular camera which is placed in MAV belly in forward looking/ Downward looking mode. The Binary Robust Invariant Scalable Key points (BRISK) algorithm was implemented for detecting feature points in the consecutive images. After robust feature point detection, efficiently performing Image Registration between the aerial images captured by MAV and with the Geo referenced images is the prime and core computing operation considered. Precise Image alignment is implemented by accurately estimating Homography matrix. In order to accurately estimate Homography matrix which consists of 9 parameters, this algorithm solves the problem in a Least Square Optimization way. Therefore, this framework can be integrated with visual odometery pipeline; this gives the advantage of reducing the computational burden on the hardware. The system can still perform the task of target geo-localization efficiently based on visual features and geo referenced reference images of the scene which makes this solution to be found as cost effective, easily implementable with robustness in the output. The hardware implementation of MAV along with this dedicated system which can do the proposed work to find the target coordinates is completed. The main application of this work is in search and rescue operations in real time scenario. The methodology was analyzed and executed in MATLAB before implementing real time on the developed platform. Finally, three case studies with different advantages derived from the proposed framework are represented.**

Keywords—Geo-localization; autonomous navigation; Visual odometery; BRISK; Homography Estimation

I. INTRODUCTION

Globally, Aerial robotics grabs the attention of the various departments starting from defense, civilian missions, photography, to video surveillance. Specifically, unmanned aerial system can address large number of problems persisting in this world .UAVs are proving to be the best solution when it comes to surveillance and reconnaissance. But still it is challenging when it comes to GPS denied unknown environments. Out of many approaches, vision based navigation is one of the trending topic because of its low cost and user friendliness. One of the major requirements in vision based navigation is to find the ground target location in terms of world coordinates which is a complex issue faced during the search and rescue operations in GPS denied environment. There are many ways to address this target Geo-localization problems [1,2,3]. Of which, this paper deals with the vision odometery based approach for estimating the ground target parameters .Some of the classic methods used optical flow algorithm for autonomous navigation which uses passive sensors like FLIR to detect and avoid obstacles but this is limited to certain range of operation.

Most of the target geo-location system uses GPS to find the location of the target. But the GPS signals are not available in the indoor as well as in urban environment due to the presence of tall structures reflect the signals from reaching the receivers which makes it less reliable. Another technique to geo locate target in the GPS denied environments is to use low MEMS INS to compute the body orientation to perform coordinate transformation as described in [1]. It may not be the efficient method due to the inherent IMU drift property. Most of the target geo location system either use active sensor like laser scanner to estimate depth or will use complex Satellite data which is very costly and not available to all.

Another important and primary step in feature based target Geo-location system is the need of robust and fast feature detection. The Scale invariant feature transform is widely known method for tracking the pixels of the interest points with respect to time. But this method will be slow in processing which lead to introduction of Speed up robust features (SURF). While the SIFT use laplacian of Gaussian for the approximation of scale space while the SURF uses the same with box filter. But when it comes to target geo location from MAV, both of them failed to give robust results in terms of error in key point matching. We used the new feature descriptor which is called BRISK [4] which can efficiently detect interest points from multiple images. BRISK reduces the

ISBN: 978−81−925974−3−0

computational cost as well as can provide optimized performance in feature point detection and matching [5]. Another significant concept of localizing MAV when flying in GPS denied environment is also addressed in this proposed framework.

The paper is organized as follows: First, we introduce the target geo location problem in GPS denied environment. Then the proposed framework and its implementation is explained in the section 3 and 4.The last section is dedicated to the test results in real time scenario with three different cases studies illustrating the advantages of the proposed frame work.

II. MOTIVATION

Still target Geo-localization from the UAV test bed with low cost is emerging one which needs a lot of development. The main motivation behind the proposed work is its application in search and rescue operation. This system can help mission operator to search and rescue any individual. It can provide valuable information on location [5] of a suspected safe house to the mission operator. Also this proposed system completely reduces computational burden by using only the information which is required for the autonomous navigation. Also the cost of the processor board is very economical when compared to the others. Nowadays, the resource for Global positioning system provided by developed countries which in turn provoke the fact of single point failure. Hence, the proposed system is very much helpful in surveillance missions. In this sort of mission the UAV should be a rucksack model. It should be dismantled and assembled in very short duration. To design and develop such a target Geo-localization system will have so many advantages for the user. The points mentioned above were the main motivation behind this proposed work.

III. PROPOSED SYSTEM ARCHITECHTURE

The proposed system consists of custom build quad copter with beagle bone as its onboard vision computer. This beagle board black is a low power consuming on board computer. This system can be used for high level computation and all the processed output can be viewed by the user in the Ground control station (GCS).The mission starts with aerial vehicle taking off from the desired location to capture the target of interest. This is done by autonomously navigating to the required location. There is an option of waypoint navigation to follow particular trajectory is available, for which latitude longitude, altitude information is given as inputs to the vehicle by the ground operator in the ground control station. The vehicle captures the information and performs an image based alignment as shown in the Fig.1. Note that the reference images are geo referenced to world coordinates which are aligned with the Geo referenced images. Our proposed framework uses the previously recorded flight images and video from same or different MAV which are geo rectified manually which acts as reference frame. By doing so, our method can be very much cost effective solution.

A. *Test bed configuration*

It consists of 4 rotors with a single diagonal arm spanning to 650mm and weighs around 2.8 kilogram which is built using carbon fibers' and other composites. It has the capability of autonomous operations using Vision as the major source of information. The system also has low quality GPS sensor which will be triggered at time of control signal lost condition between the operator and vehicle to switch over to altitude hold mode. And it can also perform blind line of sight (BLOS) autonomous take-off and landing tested using vision based guidance and control. The power system here we used was 4s 5000mAh Li-Po. It has custom built brush-less DC motor, carbon propeller for low noise. The Endurance is around 30 minutes with day/night Camera as payload. The Video telemetry was 1.2GHz 300mW (Low power) video transmitter with good ground station Antenna, delivers very good video to the ground station up to 4 miles in line of sight(LOS). It is easy to carry with a specialization of Zero tool assembly of airframe and single person back pack operation.

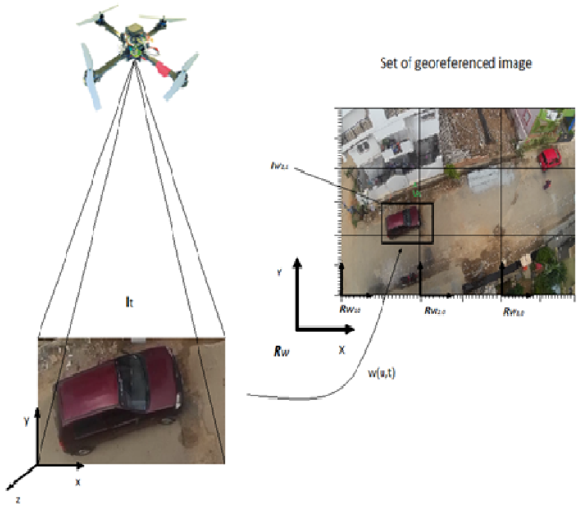

Fig.1. concept of the proposed framework

IV. PROPOSED FRAMEWORK & IMPLEMENTATION

There are mainly two major steps involved in estimating the world coordinates of the ground target in the framework

- Step 1 : BRISK Algorithm Implementation- to detect and calculate descriptor for the interest points in the aerial images taken from multiple views

- Step 2 : Perspective Transformation Estimation- to calculate transformation of the pixels from one image to another image in term of Rotation, Translation

A. *Step 1*

The BRISK ALGORTHIM is very well proven real time

Fig.1. concept of the proposed framework

ISBN: 978−81−925974−3−0

concept for detecting feature points along with the descriptors. The algorithm was applied on to the reference image and the input image coming from the MAV, to compute interest points and their correspondence in the two images. The BRISK algorithm was implemented in MATLAB.

This algorithm involves binary descriptor using which it compares the values of Gaussian pairs near to the key point. Then the algorithm follows the typical calculation of translational and rotational parameters (here it is pixels of consecutive images).

B. *Step 2*

The task of calculating the transformation is tedious involved task because if the transformations are incorrectly estimated then that will have direct effect on the final image registration between the reference and input image. After all, image registration is the task of performing warping from one frame to another based the motion parameters i.e. rotation and translation. As stated earlier, the proposed target geo location solution is part of visual odometery framework, which can be Filter based SLAM [6], feature based SLAM [7], Direct SLAM [8]. Hence the initialization part of the proposed framework uses similar SLAM methodology which is not explained here. The core part of proposed work is calculating the motion transformation between the two images which will be explained in the following.

Assumption made:

- Consider any two instants of time in which MAV is observing same world point from multiple views image at t_1 and t_2
- The world is considered as planar surface which is valid assumption because when the UAV is flying at an altitude of 50-100 meters, height variation in the ground surface is comparatively low.
- The effect of vibration on the platform is not considered.
- There are no lambertian surface in the scene
- The optical centre of the camera will never pass through the image plane.

So we have the interest point's correspondence between two images from the BRISK Algorithm implementation. The feature points detected corresponds to 3D world point when back project from image frame to world frame. Let us consider that a point P which is viewed from two different positions that is camera frame 1 and frame 2 with origin O_1 and O_2 respectively. The coordinates of point p from camera frame 1 and frame 2 is given by X_1 and X_2 respectively is given by

$$X_2 = R\,X_1 + T \quad (1)$$

The image coordinates of the same point P in image frame is given by x_1 and x_2 after some arbitrary camera rotation and translation. Refer Fig.2.

As per the assumption 2, the point p lies in the plane P, p ε P. So for any plane lying in the 3D world is characterized by a normal vector N^T. Suppose a camera is viewing a point on the plane with the offset "d" from the camera optical centre "o", then any point of X_1 on the plane satisfies this condition, $N^T X_1 = d$

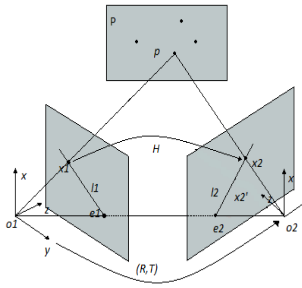

Fig.2. Multiple View Representation of point p

The component of X_1 in normal direction is the offset"d" between the plane and the camera centre.

$$\frac{N^T}{d} X_1 = 1 \quad (2)$$

Substitute (2) in (1),

$$X_2 = RX_1 + T\left[\frac{N^T}{d} X_1\right]$$

$$X_2 = \left(R + \frac{1}{d} T N^T\right) X_1 = H\,X_1 \quad (3)$$

$H = R + \frac{1}{d} T N^T \quad \epsilon\, R^{3X3}$ Where is called Homography matrix.

Hence using Homography matrix, the point in the first frame can be expressed as H times the coordinate of a same point in the 2nd frame as illustrated in Fig.1. . As Homography matrix contains rotation ε R^3, normal vector ε R and ratio of Translation to offset ε R which adds up to 9 parameters. Therefore, these parameters embraced by the Homography matrix should be estimated accurately to perform image registration which is rotation, scale and translation invariant.

Consider a plane equation

$$aX + bY + cZ = 1 \quad (4)$$

which can be rewritten in $N^T X_1 = d$ form

where the surface of a plane is given by $N^T = [a \quad b \quad c]^T$; the observed 3D world coordinate of a point is given by $X_1 = [X \quad Y \quad Z]$; $\frac{T}{d} = 1$ up to scale which is the scale ambiguity issue which is always present in 3D reconstructions.

$$[a \quad b \quad c]\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = 1 \quad (5)$$

Insert (5) *in* (1)

$$\begin{bmatrix} X' \\ Y' \\ X' \end{bmatrix} = R\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T[\,a \quad b \quad c\,]\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

ISBN: 978−81−925974−3−0

After rearranging

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = A \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (6)$$

Where $A = R + T[\,a \quad b \quad c\,]$ $\quad A \epsilon R^{3X3}$

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\left.\begin{aligned} X' &= a_1X + a_2Y + a_3Z \\ Y' &= a_4X + a_5Y + a_6Z \\ Z' &= a_7X + a_8Y + a_9Z \end{aligned}\right\} \quad (7)$$

Here $X'\,Y'Z'$ *and* $X\;Y\;Z$ are 3D world coordinates of a point p when observed from different camera locations and they are unknowns in this particular problem. In order to solve above equation, the perspective transformation condition is used. We know that the perspective transformation between world coordinate point p to image coordinates is given by

$x' = \frac{X'}{Z'}$; $y' = \frac{Y'}{Z'}$ with focal length = -1 (8)

Substituting equation (8) in (7),

$$x' = \frac{a_1X + a_2Y + a_3Z}{a_7X + a_8Y + a_9Z} \;;\; y' = \frac{a_4X + a_5Y + a_6Z}{a_7X + a_8Y + a_9Z}$$

$$x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + a_9} \;;\; y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + a_9} \quad (9)$$

Note that all terms in the above equations are known since they are image coordinates which are obtained from BRISK algorithm which have detected the interest point along with 2D coordinates of those interest points in two image frame 1 and frame 2 .Due to inherent scale ambiguity, assume $a_9 = 1 = \frac{T}{d}$, Equation (9) becomes

$$X' = \frac{AX + b}{C^TX + 1} \quad (10)$$

where $X' = \begin{bmatrix} x' \\ y' \end{bmatrix}; X = \begin{bmatrix} x \\ y \end{bmatrix} \; A = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix}; b = \begin{bmatrix} a_3 \\ a_6 \end{bmatrix}; c = \begin{bmatrix} a_7 \\ a_8 \end{bmatrix}$

An Affine transformation is special case of X' , and is obtained when c = 0. Hence by applying linear least square technique, we can solve for A, b, c matrix to obtain Homography matrix. The equation (9) can be written and arranged to obtain in the following form

$$a_7xx' + a_8yx' + x' = a_1x + a_2y + a_3;$$
$$a_7xy' + a_8yy' + y' = a_4x + a_5y + a_6$$

Hence each point like "p" in the plane gives two equations. Therefore using a minimum of 4 points pairs, we can generate 8 equations with 8 unknowns which are easily solvable using linear algebra techniques.

$$a_1x + a_2y + a_3 - a_7xx' - a_8yx' = x';$$
$$a_4x + a_5y + a_6 - a_7xy' - a_8yy' = y'$$

Rewriting the above equation in matrix form,

$$A\,X = b; \quad (11)$$

The above equation is a non homogenous Linear equation which can be solved by performing pseudo inverse on A that is by taking A^T on both sides of equation (11),

$$A^TA\,X = A^Tb\,;$$

Therefore the Homography matrix X is given by

$$X = (A^TA)^{-1}\,A^Tb \quad (12)$$

The goal of geo localizing a target from image captured from the MAV can be achieved by estimating the homography matrix more precisely. We have the geo referenced reference frame in which pixel points are expressed with corresponding 3D world coordinates. The input images from MAV with unknown frame coordinates can be geo-referenced with reference image by performing a precise image registration leading to creation of image Mosaics. The precise alignment is performed by warping function $w(x_1, \pi)$ of the input image on to reference images using the calculated Homography matrix parameters consisting of the rotation, normal vector and the ratio of translation to offset. The method to decouple these parameters from Homography matrix is beyond the scope of this work. Nevertheless, the visual odometery methods will definitely have this decoupling step in the pipeline, in order to recovery these parameters to solve structure from motion problem efficiently.

C. *Special case of the proposed algorithm*

The proposed algorithm can be used to geo localize MAV when flying in the GPS denied environment. The novelty is that, by solving for the scale ambiguity problem mentioned earlier which is inherently present in the 3D reconstruction problem, it is possible to localize MAV position accurately in locations where GPS is not present but good in features points. A low cost gimbal camera is also used for such an application proving its purpose in such a scenario. The gimbaled camera is responsible to align the camera optical axis always perpendicular to Ground plane which means that camera's image plane and earth surface (which is assumed as plane) are parallel. Most of the research work neglect scale estimation step but the proposed work considers the open sources framework developed by [9] and same was implemented in the Matlab. This scale estimation technique is well proven and easy to use and fits exactly to our framework, is the reason behind the adaptation to that particular framework. Moreover the scale estimation methodology was created for Multirotor MAVs which uses IMU and air pressure sensor. These two sensors are very well present in our custom made MAV system. By calculating the offset from the camera centre to ground plane it is possible to project the camera location in the body frame to corresponding MAV position on the ground plane which is given by

$$(X, Y, Z)_{world} = \lambda\,(X, Y, Z)_{camera} \quad (13)$$

Where λ is given by Scale parameter. The warp function of a point x_1 to transfer the same point to another coordinate is given by

$$w(x_1, \pi) = Hx_1 \quad (14)$$

Where π is the transformation (Homography matrix). The warping function contains a scale, rotation and translation of the pixel from one frame to another and is given by

$$w(x_1, \pi) = \lambda R x_1 + T \quad (15)$$

V. EXPERIMENTAL RESULTS

Real time testing using custom build aerial vehicle along with onboard target geo-localization system which consists of Beagle bone black system[10] with algorithm for estimating Homography matrix to calculate ground coordinate geo location is illustrated here. The system is designed in such a way that first 10 frames are considered for processing. Then again 10 following frames of data are not considered, in order to reduce the computational load on the Beagle Bone processor. The prime goal in this proposal was to create real time simple system to find target coordinates which is the reason for using less complex algorithm for Homography estimation with a counter balance by choosing a robust feature extraction and description technique, BRISK which is fairly fast and needs good computational resources. So some heuristic method should be chosen to equally distribute the computational resources when processing frames. Our proposed system is easily executable using most COTS products leading to providing a low cost solution for target geo location application.

Testing was conducted at different sub urban locations where GPS signal is not available. The aerial image shown in this section below as input image is one such example which was captured in between tall structures where GPS availability is not strong. The scenario is such that the MAV has gained sufficient altitude to conduct a mission for geo locating target of interest in GPS denied environments. At the same time, the reference images which can be from previous aerial footage from the different MAV or from same MAV which has captured the optimal coverage of the same sub urban area is available. This is very much cost effective technique than acquiring Satellite data from the Agencies which are definitely of high cost. At the same time, there may be security restrictions to data availability for everyone. So prior flight video data will exhaustively provide the cost effective advantage along with option to reuse the aerial video of the sub urban area captured sometime ago rather than stacking it on the hard disks for future record and analysis purpose only.

The main component in this result section is to highlight the Geo registration accuracy to locate the target in the scene. Since the proposed algorithm is depicted to be initial part of multiple view reconstruction technique leading to visual SLAM, hence accuracy of 3D reconstructions of the buildings may not be found in the results. The success of the algorithm can be observed on geo locating the target on the planar surface like roads which is shown in the results. As mentioned earlier, that is valid assumption since the MAV fly at sufficient high altitudes at which earth surface appears to plane. In order to illustrate the proposed algorithm, the MAV climbed to around 13 meters from the ground level to localize the target in the scene. The benefits of the proposed framework can be illustrated in three different cases. In first two cases, to Geo locate a new target in the scene which was present and not present in the reference scene are dealt separately. The advantage of localizing MAV based on the reference frame in GPS denied environment is depicted as third case study.

Case (i) when the target's coordinates is present within the reference frame

The target in the reference frame (in this case is Red car) which was present in both reference and input image. Note a black bounding box on the red car showing successful alignment of the input image with reference image.

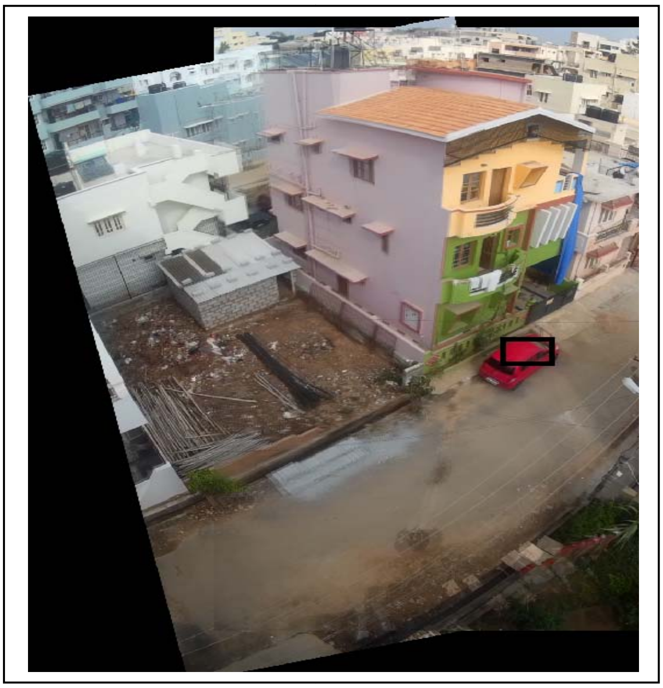

Fig.3. the processed output depicting case study 1 condition

ISBN: 978−81−925974−3−0

Case(ii) when the target's coordinates is not within the reference frame.
The target in the reference frame in this case is maroon car which is not present in the reference frame but was present in the input frame. Refer Fig.4.

Case (iii) Geo Localizing the MAV position at any point of time
The target in this case is MAV itself which must be localized to world coordinates. It was found that the scale parameter was approximately 10.9 m against the true value of 13.1m

It can be very clearly concluded that best registration (for planar surface) can be obtained provided by exactly estimation of Homography matrix. In contrast, the buildings which are not plane surfaces when observed from this altitude from MAV and in this view point, are therefore complex structures.

Fig.4. The processed output depicting case study 2 condition

They cannot be registered accurately because it is problem of structure from motion which is anyways addressed in Visual Odometery pipelines finally by solving scale ambiguity issue to obtain highly accurate reconstructions. Nevertheless, performing multiple view reconstruction is extremely challenging task [11] in real time using monocular camera.

VI. CONCLUSION

The proposed target geo-localization algorithm was successfully tested for real time images captured by custom build Mini MAV and target's world coordinates was calculated. As per the algorithm, the input images captured by the vehicle were geo-referenced to find the target's world coordinates. The images in test results section depicts that a non geo-referenced image was geo-referenced with the reference image after estimating Homography matrix precisely. The captured video was given as input to Beagle Bone Black for executing the proposed vision algorithm of estimating motion transformation between the frames. After geo-referencing the frames, the target's geo-location was readily observed since the pixels in the reference frame hold 3D world coordinates. Using previous flight images/ videos which are often stacked unused in the hard disks are used as the reference image which are manually rectified to hold world coordinates using various sources of standard reference like Open source Google map[12], prior to the mission. This gives us the cost benefits on reference image source because satellite images and private agencies data which may not be accessible to everyone and they are costly. The option to build real time system using low cost Hardware along with manual task of creating reference frames is clear advantage of this framework. All the results were obtained using implemented target geo-localization system using onboard beagle board in a real time application. The algorithm implemented behaved in same way for different real time inputs and fetched equivalent results as shown in case studies along with special case advantage of localizing MAV up to 2.2m error.

REFERENCES

[1] D.Blake Barber, Joshua D. Redding, Timothy W.McLain, Randal W.Beard,Clark N. Tylor,"Vision based Target Geo-location using a Fixed- wing Miniature Air vehicle", Journal of Intelligent Robot Systems, 2006.

[2] Clark N. Taylor,"Vision Assisted Navigation for Miniature Unmanned Aerial Vehicles (MAVs)",Final Report (Feb 2007 - Nov 2009).

[3] M.Dille, B.Grocholsky, S.Nuske, " Persistent Visual Tracking and Accurate Geo-Location of Moving Ground Targets by Small Air Vehicles",Infotech@Aerospace,AIAA,2011.

[4] S.Leutenegger,M. Chli and R.Siegwart."BRISK:Binary Robust Invariant Scalable Keypoints ", ICCV,2011

[5] Cameron Schaeffer, "A Comparison of Keypoint Descriptors in the Context of Pedestrian Detection: FREAK vs. SURF vs. BRISK," 2014.

[6] D. M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. "FastSLAM: A factored solution to the simultaneous localization and mapping problem" Proc. of the AAAI National conference on Artificial Intelligence, 2002.

[7] E. G .Klein, D. Murray, "Parallel tracking and mapping for small AR workspaces." Intl. Symp. on Mixed and Augmented Reality (ISMAR), 2007.

ISBN: 978−81−925974−3−0

[8] F. J .Engel, J. Sturm, D. Cremers. “Semi-dense visual odometry for a monocular camera.” Intl. Conf. on Computer Vision (ICCV), 2013.

[9] A. M. Achtelik, M. Achtelik, S. Weiss, and R. Siegwart. “Onboard IMU and monocular vision based control for MAVs in unknown in- and outdoor environments”. Proc. of the International Conference on Robotics and Automation (ICRA), 2011.

[10] For Beagle Board http://beagleboard.org/BLACK

[11] B. R. Newcombe, S. Lovegrove, A. Davison,” DTAM: Dense tracking and mapping in real-time.” Intl. Conf. on Computer Vision (ICCV), 2011.

[12] For more information about google earth http://www.google.com/earth/download/ge/agree.html

ISBN: 978−81−925974−3−0