Robust Visual Simultaneous Localization and Mapping for MAV Using Smooth Variable Structure Filter

Abdelkrim Nemra, Luis M. Bergasa, Elena López, Rafael Barea, Alejandro Gómez and Álvaro Saltos

Abstract The work presented in this paper is a part of research work on autonomous navigation for Micro Aerial Vehicles (MAVs). Simultaneous Localization and Mapping (SLAM) is crucial for any task of MAV navigation. The limited payload of the MAV makes the single camera as best solution for SLAM problem. In this paper the Large Scale Dense SLAM (LSD-SLAM) pose is fused with inertial data using Smooth Variable Structure Filter which is a robust filter. Our MAV-SVSF-SLAM application is developed under Linux using Robotic Operating System (ROS) so that the code can be distributed in different nodes, to be used in other applications of guidance, control and navigation. The proposed approach is validated first in simulation, then experimentally using the Bebop Quadrotor in indoor and outdoor environment and good results have been obtained.

Keywords Micro Aerial Vehicles · Simultaneous Localization and Mapping · Autonomous navigation · Data fusion · Robust filters · Sensor fusion · Robotic Operating System

1 Introduction

Self-localization of unmanned vehicles (UV) is still considered as a fundamental problem in autonomous vehicle navigation. The problem has been tackled in the recent past for different platforms and a number of efficient techniques were proposed as presented in Cox and Wilfong [1]. Many research works about optimal

A. Nemra
Ecole Militaire Polytechnique, Algiers, Algeria
e-mail: karim_nemra@yahoo.fr

L.M. Bergasa · E. López(✉) · R. Barea · A. Gómez · Á. Saltos
Department of Electronics, Alcalá University, Alcalá, Spain
e-mail: {bergasa,elena,barea}@depeca.uah.es

L.P. Reis et al. (eds.), *Robot 2015: Second Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 417,
DOI: 10.1007/978-3-319-27146-0_43

and robust data fusion have been presented in literature, Toledo-Moreo et al [2] Leonard and Durrant-Whyte [3], Toth [4]. More recently, driven by the interest of military applications and planetary explorations, researchers have turned their attention towards localization of vehicles moving in hostile and unknown environments [5, 6]. In these cases, a map of the environment is not available and hence a more complex simultaneous localization and mapping (SLAM) problem must be faced and solved.

Various techniques exist for robot pose estimation, including mechanical odometer, GPS, LASER, INS and cameras[7]. However in recent years the use of computer vision has always attracted researchers because of the analogy that we can do with the human tracking system.

SLAM has often been performed using a single camera, stereo cameras and recently, with the apparition of depth (RGBD) camera, RGBD-SLAM became an interesting solution.

Solve the Visual SLAM for Mini Aerial Vehicle (MAV) is a real challenge especially in GPS denied regions. Many alternatives of Aerial SLAM are implemented using LASER, Vision, and RGBD camera. However majority of these solutions cannot be used for MAV which can carry a limited payload (between 60-100 g). Thus, the best solution of the MAV- SLAM is to use a single camera.

The main contribution of this paper is to propose a robust solution for 6DoF-MAV-SLAM by fusing the pose given by the Large Scale Dense SLAM, and that given by the inertial navigation system using the Smooth Variable Structure Filter (SVSF). The algorithm is developed under Linux using ROS (Robotic Operating System) so that the code can be distributed in different nodes, to be used in other applications of control and navigation. The proposed solution should be able to: localize the MAV in indoor environment, should be robust face any photometric and geometric change and can maintain suitable pose accuracy even when the visual information is lost.

This paper is organized as follows, in section 2, previous work on single camera SLAM for MAV are presented. In section 3 process and observation model for IMU/LSD-SLAM algorithm are detailed. The Smooth Variable Structure Filter (SVSF) is presented in section 4. After that, the proposed algorithms are implemented and validated in simulation and experimentally in section 5. Finally, in section 6, conclusion and perspectives are given.

2 Single Camera SLAM for MAV

Real-time monocular Simultaneous Localization and Mapping (SLAM) have become very popular research topics. Two major reasons are (1) their use in robotics, in particular to navigate unmanned aerial vehicles (UAVs) [8, 9, 10], and (2) augmented and virtual reality applications slowly making their way into the mass-market.

2.1 Direct Methods SLAM

Direct Visual Odometry (VO) methods are very interesting for real time application by optimizing the geometry directly on the image intensities, which enables using all information in the image.

2.2 Large Scale Direct SLAM

Large-Scale Direct SLAM (LSD-SLAM) method, not only locally tracks the motion of the camera, but allows building consistent, large-scale maps of the environment. The method uses direct image alignment coupled with filtering-based estimation of semi-dense depth maps as originally proposed in [11]. The global map is represented as a pose graph consisting of keyframes as vertices with 3D similarity transforms as edges, elegantly incorporating changing scale of the environment and allowing detection and correction of the accumulated drift. The method runs in real-time on a CPU and even on a modern Smartphone [11]. The algorithm LSD SLAM consists of three major components: tracking, depth map estimation and map optimization as visualized in Figure. 1:

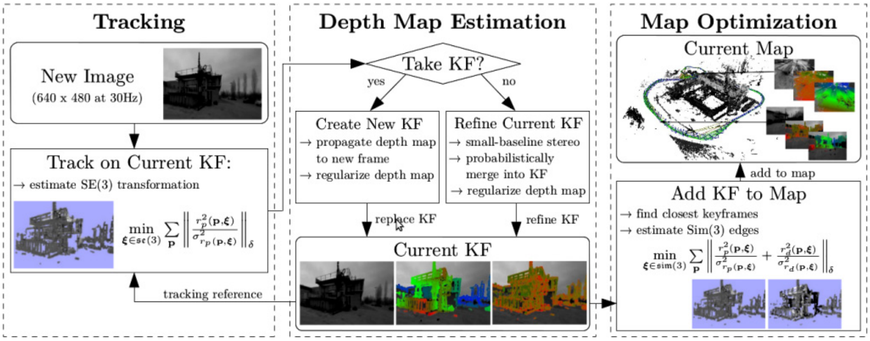

Fig. 1 Overview over the complete LSD-SLAM algorithm [11].

2.3 LSD-SLAM Limitations

Large-Scale Direct monocular SLAM (LSD-SLAM), have many advantages, firstly, it requires only a single camera, secondly, it constructs high quality map at different scales and finally and more importantly it can be run in real time on a CPU and even on a modern Smartphone. However, the LSD-SLAM algorithm suffers from several limits. Then, solving the MAV-SLAM using LSD-SLAM presents many drawbacks: - Pose and Map constructed up to scale, - very sensitive to photometric changes, - the observed scene must contain different textures and depths, - pure rotation motion cannot be estimated, - sensitive to dynamic scene. As solution, and in order to improve the performances (robustness, accuracy) of the LSD-SLAM we propose to fuse the LSD-SLAM position with inertial data using an Inertial Measurement Unit (IMU).

3 IMU/LSD-SLAM Fusion

In our work, we are using the Quadrotor Bebop which has a camera, IMU and GPS (Figure 4). For Indoor applications the GPS signal is not available.

Therefore, fuse IMU with camera data for localization and mapping is very interesting especially if the application run in real time. Furthermore, the presence of IMU data makes our MAV-SLAM able to maintain right 6DOF pose even when the visual information is not available. The proposed SLAM solution is developed under Linux using ROS (Robotic Operating System) which gives it many advantages (Parallelism, Real time, 3D map view).

3.1 Process Model (Inertial Measurement Unit)

The sate vector to be used in our SLMA problem is given by:

$$\hat{X}_{\mathrm{k}} = [x \quad y \quad z \quad u \quad v \quad w \quad q_1 \quad q_2 \quad q_3 \quad q_4]^t \tag{1}$$

Where: ($x \quad y \quad z$) the position of the UAV in the navigation frame, ($u \quad v \quad w$) : the velocities of the UAV in the body frame, ($q_1 \quad q_2 \quad q_3 \quad q_4$) : are the quaternion to represents the MAV orientation.

The cinematic model of the MAV is given by:

$$\hat{X}_{\mathrm{k}} = f\left(\hat{X}_{\mathrm{k-1}}\right) + w_k \tag{2}$$

$$\hat{X}_{\mathrm{k}} = \begin{bmatrix} p^n(k) \\ v^n(k) \\ \psi^n(k) \end{bmatrix} = \begin{bmatrix} p^n(k-1) + v^n(k)\Delta t \\ v^n(k-1) + [C_b^n(k-1) f^b(k) + g^n]\Delta t \\ \psi^n(k-1) + E_b^n(k-1)\omega^b(k)\Delta t \end{bmatrix} + w_k \tag{3}$$

w_k is the process noise, p^n, v^n and ψ^n are position velocity and orientation in the navigation frame. Where f^b and ω^b are the body-frame referenced vehicle accelerations and rotation rates which are provided by inertial sensors on the vehicle and g^n is the acceleration due to gravity. C_b^n and E_b^n are respectively the direction cosine matrix and rotation rate transformation matrix between the body and navigation frames.

3.2 Observation Model (LSD-SLAM Pose)

The position given by the LSD-SLAM will be used as observation, as follows:

$$Z_k = h\left(\hat{X}_k\right)$$

$$Z_k = \begin{bmatrix} x \\ y \\ z \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \left[\hat{X}_k\right] \tag{4}$$

As can be seen from Eq.4, the observation model is linear which is suitable for the Smooth Variable Structure Filter (SVSF) which is the subject of the next section.

4 Smooth Variable Structure Filter

While EKF-SLAM and FastSLAM are the two most popular solution methods for SLAM problem, newer alternatives, which offer much potential, have been proposed, including the use of the unscented Kalman filter (UKF) proposed by Julier and Uhlmann in SLAM [12]. Unlike the (EKF), the (UKF) uses a set of chosen samples to represent the state distribution. The UKF-SLAM avoids the calculation of the Jacobean and Hessian matrices but also obtain higher approximation accuracy with the unscented transformation (UT) [13]. However, for high-dimensional systems, the computation time is still heavy; thus, the filter converges slowly.

However, the above filters are all based on the framework of the Kalman filter (KF); it can only achieve a good performance under the assumption that the complete and exact information of the process model, observation model and noise distribution have to be a priori known.

To overcome some of these limitations, we propose to use the SVSF filter to solve the SLAM problem. Based on sliding mode concepts the first version of the variable structure filter (VSF) was introduced in 2003 [15]. On 2007, the smoothing variable structure filter (SVSF) which is relatively a new filter is introduced [14]. It is a predictor/corrector estimator based on sliding mode control theory and variable structure estimation concepts. In its old form, the SVSF is not a classical filter in the sense that it does not have a covariance matrix.

4.1 SVSF Principal

Essentially this method makes use of the variable structure theory and sliding mode concepts. It uses a switching gain to converge the estimates to within a boundary of the true state values (i.e., existence subspace shown in Figure. 2).

The SVSF has been shown to be stable and robust to modeling uncertainties and noise, when given an upper bound on the level of un-modeled dynamics and noise. The SVSF method is model based and may be applied to differentiable linear or nonlinear dynamic equations. An augmented form of the SVSF was presented in [16], which includes a full derivation for the filter using covariance matrices.

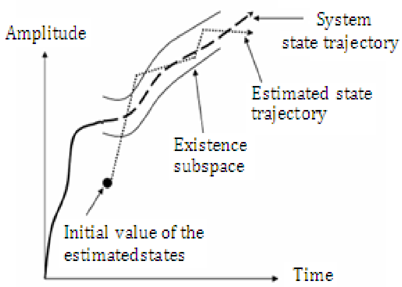

Fig. 2 The SVSF estimation concepts [15].

The basic estimation concept of the SVSF is shown in Figure 2. Some initial values of the estimated states are made based on probability distributions or designer knowledge. An area around the true system state trajectory is defined as the existence subspace. Through the use of the SVSF gain, the estimated state will be forced to within this region. Once the value enters the existence subspace, the estimated state is forced into switching along the system state trajectory.

The estimation process is iterative and may be summarized by the following set of equations (for control or estimation problem) [16]. Like the KF, the system model is used to calculate a priori state as follows:

$$X_{0/0} = X_0$$

$$E_{z,0/0} = E_{z,0}$$

$$\hat{X}_{k+1/k} = f\left(\hat{X}_{k/k}, U_k\right) \tag{5}$$

$$\hat{Z}_{k+1/k} = h(\hat{X}_{k+1/k}) \tag{6}$$

$$E_{z,k+1/k} = Z_{k+1} - \hat{Z}_{k+1/k} \tag{7}$$

$$K_{k+1} = \ diag\left[\left(\left|E_{z,k+1/k}\right|_{\mathrm{Abs}} + \gamma\left|E_{z,k/k}\right|_{\mathrm{Abs}}\right)^{\circ} sat\left(\bar{\psi}^{-1} E_{z,k+1/k}\right)\right]\left[diag(E_{z,k+1/k})\right]^{-1} \tag{8}$$

Where ° signifies Schur (or element-by-element) multiplication, the superscript + refers to the pseudo inverse of a matrix and $\bar{\psi}^{-1}$ is a diagonal matrix constructed from the smoothing boundary layer vector ψ, such that:

$$\bar{\psi}^{-1} = [Diag(\psi)]^{-1} = \begin{bmatrix} \frac{1}{\psi_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\psi_m} \end{bmatrix} \tag{9}$$

$$sat\left(\bar{\psi}^{-1} E_{z_{k+1/k}}\right) = \begin{cases} 1, & E_{zi,k+1/k}/\psi_i \geq 1 \\ E_{zi,k+1/k}/\psi_i\,, & -1 < E_{zi,k+1/k}/\psi_i < 1 \\ -1, & E_{zi,k+1/k}/\psi_i \leq -1 \end{cases} \tag{10}$$

$$\hat{X}_{k+1/k+1} = \hat{X}_{k+1/k} + K_{k+1} E_{z,k+1/k} \tag{11}$$

$$Z_{k+1/k+1} = h(\hat{X}_{k+1/k+1}) \tag{12}$$

$$E_{z,k+1/k+1} = Z_{k+1} - \hat{Z}_{k+1/k+1} \tag{13}$$

Two critical variables in this process are the a priori and a posteriori measurements error estimates, defined by (Eq.6 and 7). Note that (Eq.7) is the a posteriori measurement error estimates from the previous time step, and is used only to calculate the SVSF gain. The value of the smoothing boundary layer width vector ψ reflects the level of uncertainties and errors modelling affecting our observation model.

5 Results and Discussion

To validate the performances of the proposed approaches, we first evaluate the performances of the SVSF filter in simulation. After that, the SVSF-SLAM will be validated using the new Bebop (Figure 4).

5.1 Simulation Results

In this simulation we consider a MAV navigating using an inertial measurement unit and LSD-SLAM pose. We assume that (between $t = 600s$ and $t = 800s$) the MAV passes by a uniform region (or dark region) and no feature is detected. In other word, during this region the LSD-SLAM gives completely a wrong position (red triangle in Fig.3) with large uncertainty. As can be seen from Figure 3 when the LSD-SLAM position is lost the MAV keep follows the IMU position (blue trajectory). When t >800s the visual information return back and the LSD-SLAM pose will be fused with the IMU position again using SVSF filter.

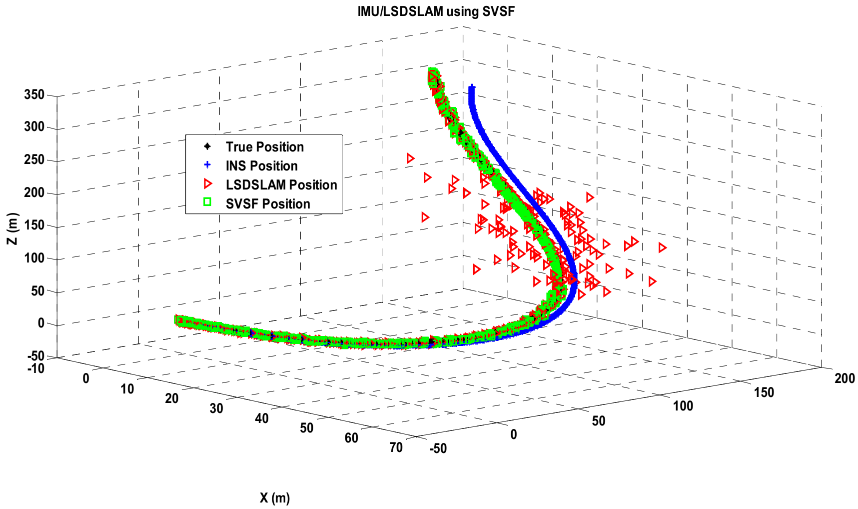

Fig. 3 Comparison of IMU, LSDVSLAM and SVSF position when visual information is not available

5.2 Experimental Results

The proposed solution of SVSF-SLAM is validated using the new MAV of parrot called Bebop (Figure 4). This latter is commercialized in December 2014 and have many advantages compared to the AR Drone2.

- **MAV Bebop Main Features**
  - Capture 1080p Video and 14MP Photos, 3-Axis Sensor (accelerometer, gyroscope and magnetometer), Linux-Based Flight Computer, Dual-core CPU, ARM 9, quad-core GPU, and 8GB of flash memory.

Fig. 4 The experimental platform the new Parrot Bebop

Experiment 1 (Indoor):
In this experiment the Quadrotor Bebop navigates in indoor environment (Figure 5). Figure 6, shows the acquired image (a), the tracked features (b) and the corresponding covariance for these features (c) using color representation. As can be seen, in this algorithm, corners and edges are selected as features, which is very important for robustness. Moreover, in the first image (c) where depth is initialized randomly then, its covariance was large (Red color), and during navigation using stereovision (Multiple view) this covariance decreases (green color).

Fig. 5 Picture of Laboratory Alcàla University

(a) (b) (c)

Fig. 6 Experimental validation of the IMU/LSD-SLAM algorithm (a) level gray image, (b) Features extraction and depth estimation, (c) Colour representation of the estimated depth variance.

Figure 7 shows the pose of the Bebop and the constructed map during navigation. The red camera represents the current position where the blue one represents the keyframe position.

Experiment 2 (Outdoor):
In this experiment the Quadrotor Bebop navigates in outdoor environment, the UAH University (Figure 8).

Figure 9 shows the pose of the Bebop and the constructed map during navigation. From Figure 9 we can observe that a good map is constructed with the MAV-SLAM as well as the followed trajectory.

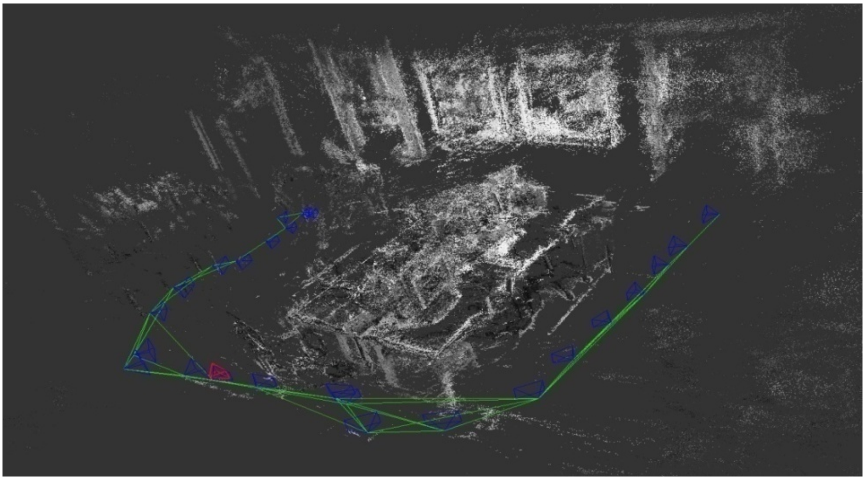

Fig. 7 Bebop localization and 3D map building using IMU/LSD-SLAM

Fig. 8 Picture of the UAH University (Alcàla)

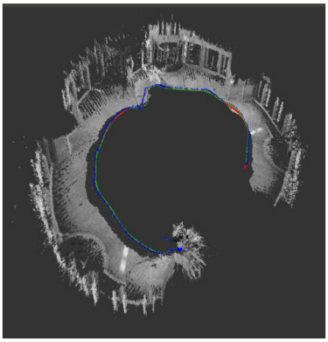

Fig. 9 Localization and 3D map building using MAV-SLAM algorithm

Experiment 3 (LSD-SLAM Pose Lost):
In this experiment which is done indoor (no GPS signal) we want to validate the robustness of our algorithm SVSF-IMU/LSD-SLAM. For that, during Bebop navigation we will hide the Bebop camera for a while (20 s) to see how the IMU/LSD-SLAM algorithm can maintain a suitable pose estimation.

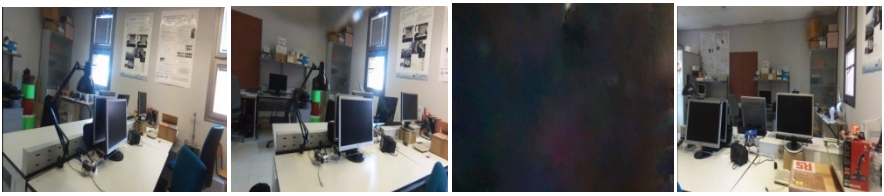

Fig. 10 Acquired image by the Bebop camera (missing of visual information for 20 s)

Figure 10 shows few acquired images by the Bebop, as can be seen some of them are completely ambiguous. Figure 11 shows the Bebop pose and the constructed map using only the LSD-SLAM algorithm, it is clear that the algorithm cannot estimate any camera (Bebop) pose when the image information is not available. Let's see what SVSF-IMU/LSD-SLAM filter can do to solve this problem. To improve the quality of the pose estimation inertial data will be used.

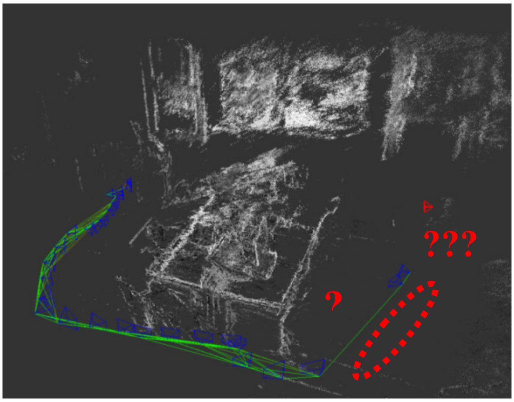

Fig. 11 Pose estimation and 3D map construction using LSD-SLAM (dashed ellipse images missing, LSD-SLAM cannot estimate any Bebop position).

Figure 12 shows the angular rates and the accelerations -following the three axis- given by the IMU during the Quadrotor navigation, when, Figure 13 shows the Bebop position given only by integrating IMU data, this latter is good for short term but it suffers from drift for long term navigation. Figure 14 shows the pose and the map of the IMU/LSD-SLAM algorithm. However because the LSD-SLAM pose is given up to unknown scale, this latter is estimated by calculating the rate between IMU-pose and LSD-SLAM-pose for few iterations in the beginning, in our case the scale factor ≈ 6. As can be seen from Figure 14, even when visual information is not available (dark region) our algorithm maintains a suitable position of the Bebop Quadrotor.

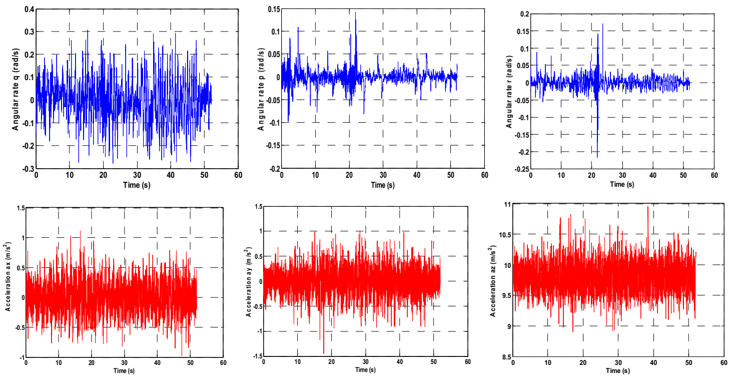

Fig. 12 Angular rates and accelerations given by the IMU

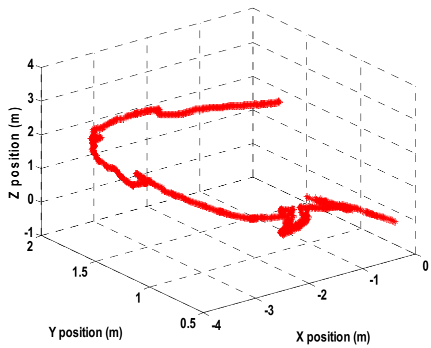

Fig. 13 Bebop position given by the IMU

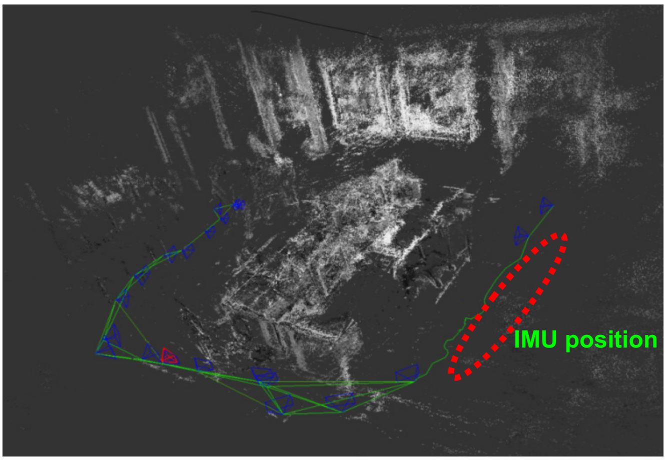

Fig. 14 Pose estimation and 3D map construction using IMU/LSD-SLAM (dashed ellipse visual information is not available, IMU/LSD-SLAM can estimate the Bebop position)

6 Conclusion

The work presented in this paper is a part of autonomous navigation for MAV. The objective was to propose an online SLAM solution for an MAV (Parrot Bebop). For that purpose, the LSD-SLAM algorithm is fused with the IMU data using SVSF filter to improve the robustness of the proposed SLAM solution, and make it able to estimate accurate pose even if the visual information is not available (images without textures, dark area...). Finally, the proposed approach is validated first in simulation, then with experiment data in indoor and outdoor environment and good results have been obtained. The next objective of our work is to implement a new version of the MAV-SLAM which should be able to estimate accurate position even in dynamic environment.

Acknowledgements This work is supported by the Index MAV Project (CCG2014/EXP-078), which is funded by the University of Alcalá, and by the RoboCity2030 III-CM project (S2013/MIT-2748) funded by Programas de I+D en la Comunidad de Madrid and co-funded by Structured Funds of the UE.

References

1. Cox, I.J., Wilfong, G.T.: Autonomous robot vehicles. Springer Verlag, New York (1990)
2. Toledo-Moreo, R., Zamora-Izquierdo, M.A., Ubeda-Minarro, B., et al.: High-integrity IMM-EKF-based road vehicle navigation with low-cost GPS/SBAS/INS. IEEE Trans. Intell. Trans. Syst. **8**(3) (2007)
3. Leonard, J.J., Durrant-Whyte, H.F.: Directed sonar sensing for mobile navigation. Kluwer Academic Publisher, Boston (1992)
4. Toth, C.: Sensor integration in airborne mapping. IEEE Trans. Instrum. Meas. **51**(6), 1367–1373 (2002)
5. Lacroix, S., Mallat, A., Chatilla, R., et al.: Rover self localization in planetry-like environments. In: 5th International Symposium on Artificial Intelligence, Robotics and Automation in Space, Noordwijk, The Netherlands, June 1999
6. Olson, C.F.: Probabilstic self-localization for mobile robots. IEEE Trans. Robot. Automat. **16**(1), 55–66 (2000)
7. Mark Theodore Draelos: The Kinect Up Close: Modifications for Short-Range Depth Imaging, Article, North Carolina State University (2012)
8. Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: fast semi-direct monocular visualn odometry. In: Intl. Conf. on Robotics and Automation (ICRA) (2014)
9. Engel, J., Sturm, J., Cremers, D.: Camera-based navigation of a low-cost quadrocopter. In: Intl. Conf. on Intelligent Robot Systems (IROS) (2012)
10. Achtelik, M., Weiss, S., Siegwart, R.: Onboard IMU and monocular vision based control for MAVs in unknown in- and outdoor environments. In: Intl. Conf. on Robotics and Automation (ICRA) (2011)
11. Engel, J., Schops, T., Cremers, D.: LSD-SLAM: Large-Scale Direct Monocular SLAM. ECCV (2014)

12. Yan, X., Zhao, C., Xiao, J.: A novel fast-SLAM algorithm based on iterated unscented Kalman filter. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO 2011), pp. 1906–1911 (2011)
13. Nemra, A., Aouf, N.: Robust INS/GPS Sensor Fusion for UAV Localization Using SDRE Nonlinear Filtering. IEEE Sensors Journal **10**(4), April 2010
14. Habibi, S.R.: The Smooth Variable Structure Filter. Proceedings of the IEEE **95**(5), 1026–1059 (2007)
15. Habibi, S.R., Burton, R.: The Variable Structure Filter. Journal of Dynamic Systems, Measurement, and Control (ASME) **125**, 287–293 (2003)
16. Gadsden, S.A., Habibi, S.R.: A new form of the smooth variable structure filter with a covariance derivation. In: IEEE Conference on Decision and Control, Atlanta, Georgia (2010)