SWIR Camera-Based Localization and Mapping in Challenging Environments

Viachaslau Kachurka, David Roussel, Hicham Hadj-Abdelkader, Fabien Bonardi, Jean-Yves Didier, Samia Bouchafa

SWIR Camera-Based Localization and Mapping in Challenging Environments

Viachaslau Kachurka✉[0000−0003−2088−6067], David Roussel[0000−0002−1839−0831], Hicham Hadj-Abdelkader[0000−0001−9944−4602], Fabien Bonardi[0000−0002−3555−7306], Jean-Yves Didier[0000−0002−9863−5471], and Samia Bouchafa[0000−0002−2860−8128]

IBISC, Univ Evry, Université Paris-Saclay, 91025, Evry, France
{viachaslau.kachurka, david.roussel, hicham.hadjabdelkader, fabien.bonardi, jeanyves.didier, samia.bouchafa}@ibisc.univ-evry.fr
http://www.ibisc.univ-evry.fr

Abstract. This paper assesses a monocular localization system for complex scenes. The system is carried by a moving agent in a complex environment (smoke, darkness, indoor-outdoor transitions). We show how using a short-wave infrared camera (SWIR) with a potential lighting source is a good compromise that allows to make just a slight adaptation of classical simultaneous localization and mapping (SLAM) techniques. This choice made it possible to obtain relevant features from SWIR images and also to limit tracking failures due to the lack of key points in such challenging environments. In addition, we propose a tracking failure recovery strategy in order to allow tracking re-initialization with or without the use of other sensors. Our localization system is validated using real datasets generated from a moving SWIR-camera in indoor environment. Obtained results are promising, and lead us to consider the integration of our mono-SLAM in a complete localization chain including a data fusion process from several sensors.

Keywords: Visual SLAM · Visual Odometry · Short-wave Infrared (SWIR) camera.

1 Introduction

The problem of accurate localization of emergency response agents (civil security, firefighters, etc.), law enforcement or armed forces agents in a closed, unknown, non-cooperative environment remains an open problem nowadays since no sufficiently reliable system meeting all specific constraints currently exists. However, many military and civil applications would benefit from being equipped with such systems. Such localization task focuses on the idea that the command center should have the most accurate location of its agent in unknown conditions, while also receiving information about the environment (e.g. reckon missions in armed forces, or operative information on a fire).

While indoor positioning problem by itself already imposes additional difficulties, such as a need of high accuracy level and non-existence of GPS signal [12], the given problem formulates even a higher-level difficulty extension to it.

One can see such task as a use case for Simultaneous Localization and Mapping (SLAM) techniques; however, the main challenge for SLAM techniques in such context is the lack of suitable technologies that can take into account the technical limits (the equipment should be quite small and efficient), technological requirements (diversity of sensors to make the system more robust to ensure the mission) and environmental constraints (hazardous or non-cooperative environment), as shown in the article which defines a similar problem [18].

The multi-sensor solutions have been studied in the field of mobile robotics for several decades and usually consider a wheeled vehicle, moving in pretty homogeneous conditions: such as a mobile robot in [16] or, more recently, a flying drone [6]. In most cases data fusion from multiple sensors is required with special interest in combination of inertial measurement unit (IMU) with other sensors such as cameras [17] or LIDAR [11].

While during decades the imaging sensor appeared to be among the least appealing in the field of robust real-time indoor positioning due to high computational complexity and susceptibility to fail in non-cooperative environment, recent convergence of visual SLAM field (as observed in [21]) enabled more robust approaches and reopened this niche. Consequently, this last decade has seen a profusion of works in the field of localization and SLAM. However, there exists a certain lack of works offering hardware and software solutions specific to complex environments.

The aforementioned "non-cooperative environment", as described in [18], is an environment where the conditions tend to render the work of any type of localization approach as difficult as possible. In the context of a visual odometry approach, some relevant constraints that should be addressed, are: rapid and drastic change of light conditions such as outdoor / indoor transitions, presence of heavy smoke and human motion which implies a wide range of irregular translations and rotations speeds. In the frame of this work we solely focus on the aforementioned visual odometry problem under major limitations of the imaging sensor, trying to assess applicability of some existing approaches to this ambitious task, as a part of a larger multi-sensor system which is out of scope of this paper. Also, among the various constraints we have taken into account, the current localization should be available (and possibly transmitted to control center) in real-time, whereas complete localization trajectory and reconstructed map can be retrieved and processed later.

This paper is organized as follows: next section is devoted to the specific short-wave infrared imaging system that we propose in order to take into account some complex smoky environments. A study of the sensor spectral characteristics along with the most suitable features that we can extract from the resulting images is provided. Section 3 shows how a classical SLAM (here ORB-SLAM) algorithm is adapted to meet our specific requirements. In particular, we focus on the tracking re-initialization step. Section 4 presents experimental results on

(a) Room viewed by standard RGB-camera (b) Room viewed by SWIR camera

Fig. 1. A medium-sized room, filled with cold smoke with detected ORB [20] features

existing benchmarks and data obtained from our shortwave infrared camera, moving in an indoor and outdoor environment. Finally, the paper ends with a conclusion and some future works.

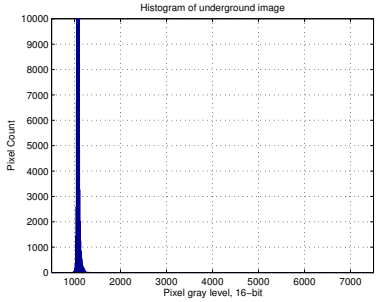
2 SWIR Camera Characteristics and Features

A conventional RGB-camera sensor does not see through heavy smoke. Fig. 1a and 1b present a room, filled with common dry smoke. Nevertheless, most types of smoke (either artificial dry smoke [22] or several types of "natural" smoke [1]) are transparent to infra-red imaging sensors, most likely for the Short-Wave Infra-Red (SWIR) camera — in the spectral band 0.9–1.7 µm.

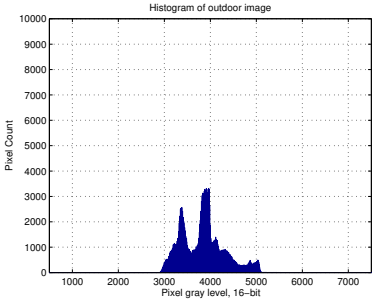
However, SWIR imaging is still subject to poor lighting conditions provided by dark places or cold lights such as fluorescent or LED lighting, since the latter ones do not emit enough light in the SWIR band [23]. An example of a SWIR image of the underground parking, lighted-up with neon tubes, is shown in Fig. 2 (c), where the scene is clearly unseen by the SWIR camera. Otherwise, this visibility limitation does not appear when the scene is illuminated by light bulb or sun (see illuminated room beyond the doorway in Fig. 2 (c) and the outdoor parking in Fig. 2 (d)). Hence, cold lighting conditions would require an external SWIR-band emitting light source to illuminate the scene (e.g. the one described in [5]).

Non-cooperative environments can feature drastic and rapid changes in the global lighting conditions. Such changes can be described by transition from dark indoor to sunny outdoor (and vice-versa), moving elements (flashlights or personnel), or non-constant lighting produced by open fire.

Also, as there can be no information of the light conditions of the explored area *a priori*, we can not apply any photometric calibration or rely on the data about scene luminosity, used in most of direct visual SLAMs. A transition from underground to sunny outdoor parking lot presents such lighting condition differences, that the resulting histograms do not even intersect within the chosen



(a) Underground histogram



(b) Outdoor histogram

(c) Underground parking lot after histogram equalization

(d) Outdoor parking lot after histogram equalization

Fig. 2. Top row: Histograms for raw image data, extracted from SWIR camera with exposure time of 20 ms. **Bottom row:** the same images after histogram equalization treatment (automatic gain control-like algorithm, AGC)

camera sensitivity spectrum (encoded with 16 bits depth) as it can be seen on Fig. 2 (a) and (b). This leads to usage of general histogram equalization approach, forbidding any direct visual SLAM such as DSO [4] (which requires an accurate photometric calibration and constant shutter time [3]). In our case we employed an automatic gain control-like algorithm (AGC), which deletes the points beyond 3σ-limits from both sides of histogram, and then spreads it across the whole 16-bit range.

Such an approach defies the problems of rapid luminosity changes, also avoiding the glare effect, and provides comparable images for comparable scenes in different luminosity conditions. However, it fails in the cases of very narrow histogram (as also can be seen in the left column of Fig. 2) and introduces noise. The only way to avoid very narrow histograms is to add a portable infrared light source as discussed previously. This line of consideration leads us to an idea of using an indirect visual SLAM, which relies on a feature detector algorithm.

2.1 Feature Detection in Infrared Imaging Sensors

While there has been a lot of research in recent years regarding feature points, few of them concern infra-red sensors. These sensors usually provide images with characteristics not equal to those of standard cameras. Therefore, one cannot automatically apply their most known *pros et cons* to the task of feature detection and description in IR-imaging.

Most known researches in this particular direction were Ricaurte et al. [19], which compared the feature detection and description efficiency against several typical image transformations on the long-wave infrared (LWIR) imaging sensors. Johannson et al. propose in [10] a similar work for IR-images. Indeed, they consider the ORB [20] detector and descriptor couple as the second-best regarding robustness and efficiency, losing only to the combination of ORB detector and BRISK descriptor.

Therefore, one of the best compromises, combining these results with the overall time efficiency [13], is a visual SLAM, based on ORB detector-descriptor, such as ORB-SLAM, introduced in [14] by Mur-Artal et al.

3 Shortwave Infrared Monocular ORB-SLAM

One of the strongest points of ORB-SLAM according to the comparative study in [9] is the usage of the same ORB descriptors for tracking, map point generation, and environment recognition. This enables a bag-of-words (BoW) based scene description [7], and therefore a fast relocalization. However, this approach also bases itself on the assumption that the movement between two consecutive frames is relatively small, which is not always the case in the context of a human agent in non-cooperative environment (rapid turns, pose changes, etc).

3.1 ORB-SLAM: Short Technical Description

ORB-SLAM bases itself on the idea of KeyFrames observing MapPoints (generated from matched features, observed during three consecutive KeyFrames). The consecutive KeyFrames are bound into an "essential graph", sub-graph of a "covisibility graph", where the KeyFrames are connected if they both observe a significant number of common MapPoints.

As most contemporary visual SLAMs, ORB-SLAM employs tracking, local mapping, loop closure, as well as bundle adjustments, both global (GBA) and local (LBA), in a multi-threaded framework. It also introduces a novel approach to monocular initialization, based on random sample consensus (RANSAC), which usually catches the movement within 10–15 frames and initializes the tracking and mapping with point and trajectory positioning up to a scale factor. In the context of our task, where the visual odometry is seen as a part of a bigger multi-sensor system, the problem of scale factor should be addressed on a higher level of a multi-sensor fusion.

GBA is employed only in the cases of LC and relocalization as it consumes a significant amount of resources. Both loop closure and relocalization work in

a similar manner - comparison of the scene BoW signature with those "already seen", and relocation of current camera position to an already existing matched scene, with further propagation of error between these two positions along the whole trajectory via GBA. Relocalization allows to resume tracking when it fails.

Such tracking failures can be pretty numerous due to the assumption of small movement between frames: if current frame and the last KeyFrame have a decreasing number of mutual MapPoint observations, a new KeyFrame is created. However, if this decay is too fast, tracking might be lost. The next section addresses this specific issue.

3.2 Tracking Re-initialization

Most SLAM algorithms are designed to work on existing benchmark datasets such as KITTI VO [8] or EUROC MAV [2], and as such are tailored not to fail on such datasets. This is often not the case when we submit these algorithms to more difficult conditions in which visual tracking can fail. This type of tracking loss occurs, for example, during fast rotations or high acceleration motions which are likely to occur when the tracking system is worn by a human being.

The default behavior of the visual tracking when the tracking is lost leads to a relocation based on BoW-signatures. However, it is likely to succeed only when the camera returns to a location already registered in a KeyFrame, which might take a while to occur and therefore induce a gap in localization data. We have therefore modified the default behavior to initiate a nondestructive tracking reset (thus preserving the KeyFrames and MapPoints recorded in the current Map) in parallel with the relocalization procedure.

Fig. 3 shows the scheme of tracking algorithm, divided into several states and procedures, where the relocalization was the only usable procedure by default when tracking was lost. We added a new state "REINITIALIZING" and a new procedure of re-initialization to restart the tracking based on the current motion, without having to wait for a possible relocalization, while preserving the current map. This procedure initializes new tracking with a new map (then merged with the old map) from a given initial pose. In standalone mode the re-initialization procedure uses the last motion, available before the loss of tracking, to provide such an initial pose.

However, multi-sensor-based SLAM can provide the most accurate possible initial pose when the re-initialization procedure succeeds, and fixes the scale factor problem of pure monocular tracking. We should also mention, that a visual-inertial extension of ORB-SLAM has already been presented in [15], however without an open-source implementation it can not yet be assessed and used *as is*.

4 Experimental Results

The task of visual odometry in last years has been very popular in the field of robotic navigation and even has grown to have a competition against several

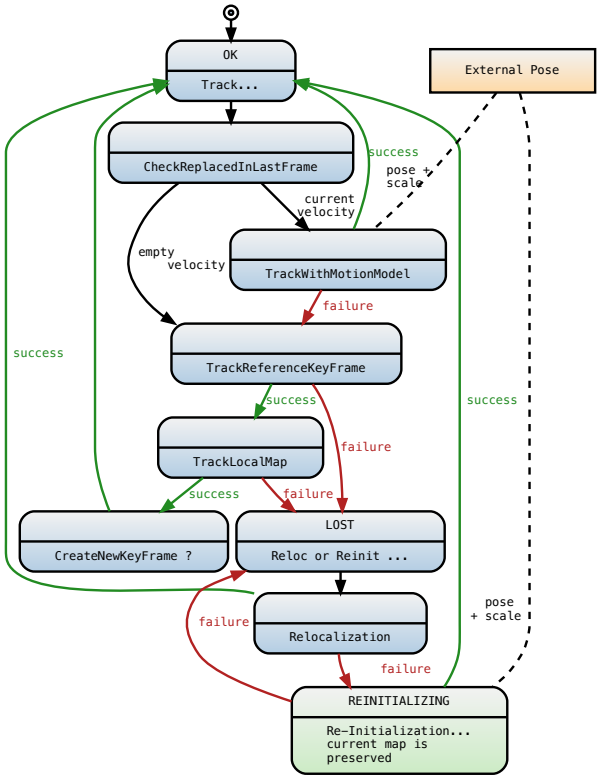


Fig. 3. Part of the general ORB-SLAM tracking scheme following a successful initialization, showing the tracking states "OK" and "LOST", as well as the tracking stages.

renown data benchmarks. One can cite the famous ones like the aforementioned EUROC MAV, KITTI VO, as well as TUM MVO [3].

EUROC MAV and KITTI VO datasets are based on camera motion, mounted on a vehicle, featuring less pitch and roll compared to human motion, and therefore do not meet all the criteria of our operating context.

In this section, we first validate the proposed re-initialization approach using the TUM MVO dataset since it is produced by a handheld camera for tracking evaluation. Then, we validate the usage of the SWIR-camera based SLAM under several constraints.

4.1 TUM MVO Dataset

In order to be able to achieve stable initialization and tracking for the TUM MVO dataset sequences, we had to adjust several parameters of ORB-SLAM. Table 1 shows the adjusted parameters compared to the values used in origianl version for respectively Outdoor and Indoor tracking. Lowering various thresholds and increasing the RANSAC iteration count provides better grip on tracking in any

Table 1. Parameters values, used in our experiments, for more stable tracking in human movement

Parameter name	Original	Outdoors	Indoors
Number of ORB extractor features	1000	2000	2000
Initial FAST threshold	20	15	20
Minimal FAST threshold	7	5	10
ORB matcher lower threshold count	50	40	40
Minimal projection matches number	20	15	15
Minimal inliers number after reloc	50	30	30
Motion model minimal matches number	20	15	15
Initializer RANSAC iterations count	200	300	300
Initializer min matched keypoint count	100	30	30

situation. However, these modifications can diminish the quality of triangulation since the scale drift can increase. Therefore, low thresholds are suggested when other sensors can be used to correct the drift on the higher level of data fusion.

The TUM MVO dataset lacks full-trajectory ground truth (GT) data, providing only partial coverage. Therefore, in the view of re-initialization validation, we are not going to run the qualitative tests against this partial GT data (besides, the original TUM MVO paper already presents the results of such tests [3]); moreover, our target here is to test the algorithm stability against tracking failures with different configurations: "Original" (O), with original parameters values as provided by Mur-Artal and no re-initialization; "Original with Reinit" (OR), with original parameters values but re-initialization added; and "Modified with Reinit" (MR), with ajusted parameters values from Table 1 and re-initialization.

We made the system do 25 runs for each of these three sets of parameters against several chosen sequences in TUM MVO dataset, in order to count the percentage of lost data (due to tracking failures). Table 2 shows the results of such validation: for sequence 24 the original set of parameters with reinitialization works better, than ours; sequence 35 shows drastic difference, and other sequences show a small increase of efficiency. This shows a crucial need for a fine tuning (or even a strategy of on-the-fly adjustment) of the parameters in each case even for the same hardware combinations, and therefore, a necessity of an additional study.

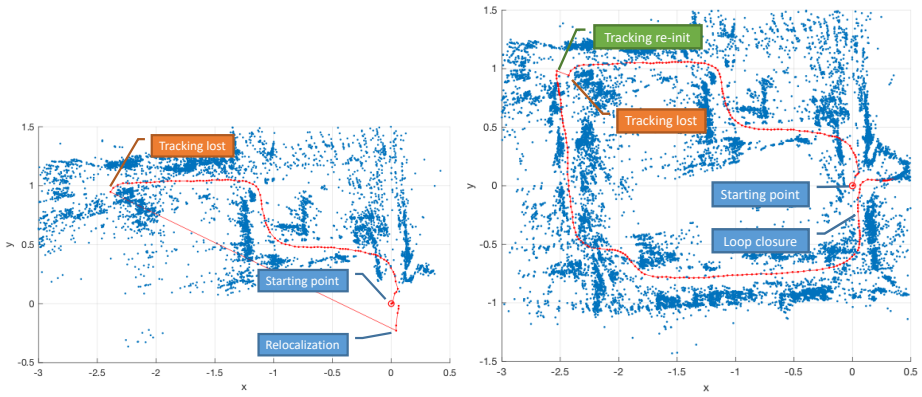
4.2 IBISC SWIR Dataset

The dataset we used in the following experiments represents a capture of a handheld SWIR camera during an exploration scenario. It is composed of about 60K images, captured at a frequency of 29 frames per second, for a duration of about half an hour. The AGC-like histogram equalization approach, mentioned in Section 2, is applied automatically to each image.

The exploration course presents multiple difficulties, which favor challenging tracking situations, such as: rapid changes of direction, indoor / outdoor

Table 2. Levels of tracking failures for "original" (O), "original with reinitialization" (OR) and "modified with reinit" (MR) configurations against several TUM MVO sequences. The percentage level shows, how much frames were lost during state "Tracking lost", as compared to total frame count of the sequence.

Sequence num.	Data loss, % (O)	Data loss, % (OR)	Data loss, % (MR)
Seq. 24 (parking)	33.93	**5.14**	5.97
Seq. 25 (parking)	6.94	1.62	**1.34**
Seq. 29 (street)	2.24	0.53	**0.00**
Seq. 30 (backyard)	48.38	19.47	**18.47**
Seq. 35 (indoors)	58.74	56.69	**30.97**



(a) Tracking failure (b) Tracking failure + re-initialization

Fig. 4. Examples of tracking with and without re-initialization

transitions, doorways crossing. It also features regular "loop closing locations" (passing through the same place several times), which allow trajectory and map optimization through global bundle adjustment.

Fig. 4 shows two tracking scenarios using the same sub-sequence of the aforementioned dataset: Fig. 4 (a) shows a tracking failure due to a fast rotation with a relocalization event at the end of the trajectory. Since ORB-SLAM uses a random sample consensus to choose points during tracking, it is quite common to see the tracking either succeed or fail on the same data. Fig. 4 (b) shows the same tracking failure followed by a tracking re-initialization, hence preserving previous trajectory and map with also a loop closure event at the end of the trajectory. Moreover, since we only used visual tracking without integrating any other sensor, the motion model used during re-initialization assumes a continuous motion estimated over the last frames before tracking failures, which can lead to inconsistent reinitialized location if the re-initialization takes too much time (more than 1 second). In our case the trajectory has been regularized by the loop closure event which triggers a global bundle adjustment along the whole trajectory. It would be appropriate during this re-initialization to use data from other

sensors (inertial unit for instance) to obtain a more consistent re-initialization pose.

5 Conclusion and Future Work

This paper has presented a first step towards a complete localization system in challenging environments. We have shown how a specific camera (SWIR) with an adaptation of a SLAM technique (ORB-SLAM) is a promising solution. Our next step will be to integrate the proposed approach in a multi-sensor system including an inertial measurement unit in a local fusion scheme. Another perspective of our work is to couple the SWIR camera with a conventional one in order to benefit from a heterogeneous stereo image pair. The main advantage of this latter solution is to provide a complete visual SLAM that is able to fix the scale factor without any fusion with other sensor. The main fusion process in this case will remain global, by combining homogeneous poses estimations from different sensors / algorithms. Eventually, we would also like to release publicly a SWIR image dataset, dedicated to non-cooperative environment. Such task is not possible for us yet, as it lacks ground truth estimation.

Acknowledgements. This work takes part in the LOCA3D project in the framework of the challenge MALIN funded with the support of Directorate General of Armaments and French National Research Agency (https://challenge-malin.fr).

References

1. Bergstrom, R.W., Pilewskie, P., Russell, P.B., Redemann, J., Bond, T.C., Quinn, P.K., Sierau, B.: Spectral absorption properties of atmospheric aerosols. Atmospheric Chemistry and Physics **7**(23), 5937–5943 (2007)
2. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research **35**(10), 1157–1163 (2016)
3. Engel, J., Usenko, V., Cremers, D.: A photometrically calibrated benchmark for monocular visual odometry. In: arXiv:1607.02555 (Jul 2016)
4. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 611–625 (Mar 2018)
5. Ettenberg, M.H., Blessinger, M.A., O'Grady, M.T., Huang, S.C., Brubaker, R.M., Cohen, M.J.: High-resolution SWIR arrays for imaging at night. In: Infrared Technology and Applications XXX. vol. 5406, pp. 46–56. International Society for Optics and Photonics (2004)
6. Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: Fast semi-direct monocular visual odometry. In: 2014 IEEE international conference on robotics and automation (ICRA). pp. 15–22. IEEE (Jun 2014)
7. Gálvez-López, D., Tardós, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics **28**(5), 1188–1197 (Oct 2012)

8. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (Aug 2013)
9. Huletski, A., Kartashov, D., Krinkin, K.: Evaluation of the modern visual SLAM methods. In: 2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT). pp. 19–25. IEEE (Nov 2015)
10. Johansson, J., Solli, M., Maki, A.: An evaluation of local feature detectors and descriptors for infrared images. In: European Conference on Computer Vision. pp. 711–723. Springer (Oct 2016)
11. Levinson, J., Thrun, S.: Robust vehicle localization in urban environments using probabilistic maps. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 4372–4378. IEEE (May 2010)
12. Mautz, R.: Overview of current indoor positioning systems. Geodezija ir kartografija **35**(1), 18–22 (2009)
13. Miksik, O., Mikolajczyk, K.: Evaluation of local detectors and descriptors for fast feature matching. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). pp. 2681–2684. IEEE (Nov 2012)
14. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE transactions on robotics **31**(5), 1147–1163 (Aug 2015)
15. Mur-Artal, R., Tardós, J.D.: Visual-inertial monocular SLAM with map reuse. IEEE Robotics and Automation Letters **2**(2), 796–803 (Jan 2017)
16. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 1, pp. I–I. IEEE (Jul 2004)
17. Nützi, G., Weiss, S., Scaramuzza, D., Siegwart, R.: Fusion of IMU and vision for absolute scale estimation in monocular SLAM. Journal of Intelligent & Robotic Systems **61**(1), 287–299 (Jan 2011)
18. Rantakokko, J., Rydell, J., Strömbäck, P., Händel, P., Callmer, J., Törnqvist, D., Gustafsson, F., Jobs, M., Grudén, M.: Accurate and reliable soldier and first responder indoor positioning: multisensor systems and cooperative localization. IEEE Wireless Communications **18**(2), 10–18 (Apr 2011)
19. Ricaurte, P., Chilán, C., Aguilera-Carrasco, C., Vintimilla, B., Sappa, A.: Feature point descriptors: Infrared and visible spectra. Sensors **14**(2), 3690–3701 (Feb 2014)
20. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: Proceedings of the 2011 International Conference on Computer Vision. pp. 2564–2571. ICCV '11, IEEE Computer Society, Washington, DC, USA (2011)
21. Saputra, M.R.U., Markham, A., Trigoni, N.: Visual SLAM and structure from motion in dynamic environments: A survey. ACM Computing Surveys (CSUR) **51**(2), 37 (Jun 2018)
22. Schneider, J., Koch, E.C., Dochnahl, A.: Method of producing a screening smoke with one-way transparency in the infrared spectrum (Nov 26 2002), US Patent 6,484,640
23. Schubert, E.F.: White light sources based on wavelength converters. In: Light Emitting Diodes, chap. 21, pp. 346–366. Cambridge University Press, 2nd edition edn. (2006)