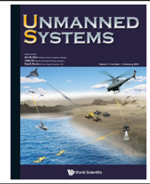
Unmanned Systems, Vol. 10, No. 3 (2022) 273–288

DOI: 10.1142/S2301385022500157

UNMANNED SYSTEMS
https://www.worldscientific.com/worldscinet/us



Multispectral Visual Odometry Using SVSF for Mobile Robot Localization

Benyounes Fahima*,‡, Nemra Abdelkrim†,§

*Laboratoire Véhicules Autonomes et Intelligents,

†Laboratoire Guidage et Navigation,
Ecole Militaire Polytechniques,
BP17, Bordj El Bahri, Algiers, Algeria

In this paper, we propose a novel method for mobile robot localization and navigation based on multispectral visual odometry (MVO). The proposed approach consists in combining visible and infrared images to localize the mobile robot under different conditions (day, night, indoor and outdoor). The depth image acquired by the Kinect sensor is very sensitive for IR luminosity, which makes it not very useful for outdoor localization. So, we propose an efficient solution for the aforementioned Kinect limitation based on three navigation modes: indoor localization based on RGB/depth images, night localization based on depth/IR images and outdoor localization using multispectral stereovision RGB/IR. For automatic selection of the appropriate navigation modes, we proposed a fuzzy logic controller based on images' energies. To overcome the limitation of the multimodal visual navigation (MMVN) especially during navigation mode switching, a smooth variable structure filter (SVSF) is implemented to fuse the MVO pose with the wheel odometry (WO) pose based on the variable structure theory. The proposed approaches are validated with success experimentally for trajectory tracking using the mobile robot (Pioneer P3-AT).

Keywords: Mobile robot localization; multispectral vision; visual odometry; stereovision; trajectory tracking; data fusion; SVSF; EKF.

1. Introduction

The major problem of mobile robotics is autonomous navigation [1,2], which means that the mobile robot must be able to build a map, locate itself within this map and perform the requested mission. So the mobile robot has to perform three basic tasks: localization, planning and navigation. Localization relative to the environment is of potential importance as it determines the smooth running of the other two tasks. It consists in calculating and maintaining the knowledge of the robot's position and orientation in an absolute reference linked to the environment work [3]. To achieve localization, a robotic system must be able to perceive the environment in which it evolves and calculate its position (estimate its position) [4,5].

Different sensors and techniques exist to estimate the position and orientation of the robot, among which are odometry [6,7], inertial system (INS), GPS and LASER [8]. However, in recent years, the use of vision in mobile robotics has always attracted researchers because of the analogy we can make with the human localization system. Localization methods based on vision [9] (SLAM, visual odometry) [10] are largely developed in recent years especially on visible images captured by one or more monocular cameras [11].

Classical visual localization uses visible images and it suffers from several problems, such as features' extraction and matching, perceptual aliasing and nighttime localization.

Many solutions are proposed in the literature to overcome these limitations; in Ref. 12, the authors have proposed an approach based on integrate point-cloud segmentation with 3D LiDAR scan-matching for mobile robot localization and mapping. Another solution has been proposed in Ref. 13 which consists in computer stereovision for

Received 15 April 2021; Accepted 6 September 2021; Published 27 November 2021. This paper was recommended for publication in its revised form by editorial board member, Zhi Gao.
Email Addresses: ‡benyounes.fahima@gmail.com, §karim_nemra@yahoo.fr

autonomous driving, where the authors talk about visual feature detection, description and matching. Xuyou Li *et al.* in Ref. 13 proposed to use 6 DoF simultaneous localization and mapping (SLAM) using the standard ICP method and its variants. A new algorithm is illustrated in Ref. 14, where the authors propose a multimodel EKF integration navigation algorithm and estimate the measurement errors of beacons online.

In this paper, a novel method for multispectral visual odometry (MVO) for mobile robot localization and navigation is proposed. Unlike the classical visual odometry [9,10], the proposed MVO consists in combining visible and infrared images to localize the mobile robot under different conditions (day, night, indoor and outdoor). Similar to the visual odometry algorithms, the MVO algorithm is based on two main steps: first, 3D landmarks are localized in the word frame using multispectral images (RGB, IR and depth images) from the Kinect sensor [15,16]. Second, constructed 3D landmarks are associated in order to localize the mobile robot.

The depth image acquired by the Kinect sensor is very sensitive to IR luminosity because the infrared sensor is parasitized by the sunrays [17], which makes it not very useful for outdoor localization. Moreover, an RGB camera requires enough lighting to be used for features' extraction during nighttime. Thus, in order to make the Kinect more useful, we propose an efficient solution for the Kinect limitation based on multimodal visual navigation [13]. The proposed solution combines the Kinect data (RGB image, IR image and depth image) [17–19] in three navigation modes: indoor localization mode based on RGB/depth images, nighttime localization mode based on depth/IR images and outdoor localization mode using multispectral stereovision RGB/IR. The latter is very important to extend the utility of the Kinect sensor in outdoor environment. For automatic selection of the appropriate navigation modes, a fuzzy logic controller (FLC) is proposed using multispectral images' energies as the input and navigation modes as the output. To improve the robustness of the proposed algorithm and to overcome the limitation of the multimodal visual navigation especially during navigation mode switching, the MVO pose is fused with the wheel odometry (WO) pose using, first, an extended Kalman Filter (EKF) and second, a smooth variable structure filter (SVSF). The latter does not make any assumption about noise characteristics and provides a robust estimation of the mobile robot pose from multiple sensors. When visual information is not available, the WO localization system will be used. The proposed approaches are validated and evaluated experimentally for a trajectory tracking problem with the mobile robot (Pioneer P3-AT). Many scenarios are considered in indoor, outdoor and during nighttime. Good results of 3D features' estimation and mobile robot localization and navigation are obtained under any conditions.

Fig. 1. Pioneer mobile robot.

The main contributions of this article are, first, the proposition of a new MVO for multimodal localization for mobile robots. Second, optimal and robust fusion of the MVO and WO pose using the EKF filter and SVSF. Finally, a FLC has been proposed for automatic selection of the suitable navigation modes (indoor, outdoor or nighttime) using images' energies. These contributions are validated in simulation and experimentally using the mobile robot P3-AT (Fig. 1).

This paper is organized as follows, first, in Sec. 2, we will talk about system modeling (process and observation model) for the mobile robot and Kinect sensor. Second, in Sec. 3, the 3D estimation using multispectral vision is described. Third, in Sec. 4, we will present the MVO for multimodal navigation (MMN) as well as the MVO algorithm based on the ICP-SURF algorithm. After that, in Sec. 5, the SVSF is presented, and then we will talk about mobile robot localization using EKF-MVO and SVSF-MVO algorithms. In Sec. 6, we present the strategy of navigation mode selection using FLC; before ending, experimental validation, results and discussion are given in Sec. 7. Finally, in Sec. 8 a conclusion is presented.

2. Mobilization

2.1. *Mobile robot model*

The mobile robot, Pioneer P3-AT (Fig. 1), is a non-holonome four-wheel robot. This kind of robot is unable to make instantaneous translations parallel to the axis of the wheel. Such a move requires many maneuvers [20].

The kinematic modeling of a mobile robot makes it possible to obtain the trajectory that it must describe by knowing its initial configuration and its translational and rotational speeds.

The mobile robot moves in 2D environment. The robot pose vector is given by $R = (x_r, y_r, \theta)^T$, with (x_r, y_r) being the coordinates of the robot in the global coordinate system and θ being the robot orientation in the global frame (Fig. 2).

June 7, 2022 12:01:09pm WSPC/284-US 2250015 ISSN: 2301-3850 2nd Reading

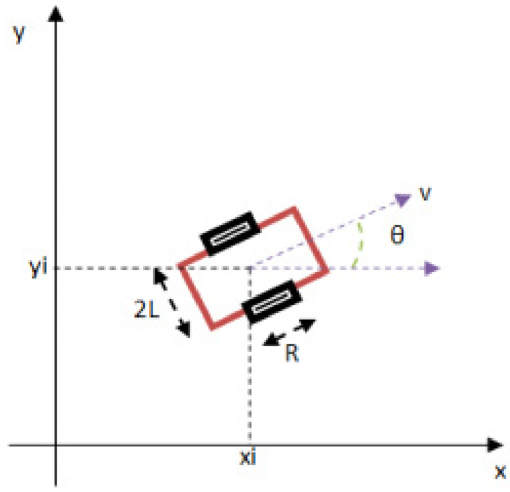


Fig. 2. Presentation of robot in the global frame.

The non-holonomic constraint is written as follows [10]:

$$y \cdot \sin(\theta) - y \cdot \cos(\theta) = 0. \tag{1}$$

From Fig. 2, we get

$$\begin{cases} \dot{x}_r = V_t^x \cos(\theta) - V_t^y \sin(\theta), \\ \dot{y}_r = V_t^x \sin(\theta) - V_t^y \cos(\theta), \\ \dot{\theta} = V_r, \end{cases} \tag{2}$$

where V_t^x and V_t^y are the robot translation speeds along the x and y axes of the local coordinate system linked to the robot, respectively, and V_r is its rotation speed.

As the robot does not move sideways, i.e., $V_t^y = 0$, so the cinematic model of the robot will be as follows:

$$\begin{cases} \dot{x}_r = V_t \cos(\theta), \\ \dot{y}_r = V_t \sin(\theta), \\ \dot{\theta} = V_r. \end{cases} \tag{3}$$

2.2. *Observation model*

A standard visual sensor collects Red, Green and Blue (RGB) wavelengths of light. Multispectral sensors are able to collect, in addition to visible wavelengths, many infrared radiation wavelengths [21]. Multispectral imaging combines many spectral bands (from 3 to 15 nm) (Fig. 3).

We have chosen the Kinect 360 V1 because it is less expensive and not greedy in computing time (Fig. 4). However, as we have already mentioned above, its operation is limited because the infrared sensor is parasitized by the sunrays; furthermore, it cannot observe objects within a range less than 500 mm. The Kinect camera principal is based on structured light; it contains two cameras, an RGB camera that provides visible specter and an IR camera which acquires near infrared (NIR) images. To provide the depth image of the 3D observed scene, a 3D LASER projector is combined with the IR camera. Stereovision between an RGB and an IR image is called multispectral stereovision.

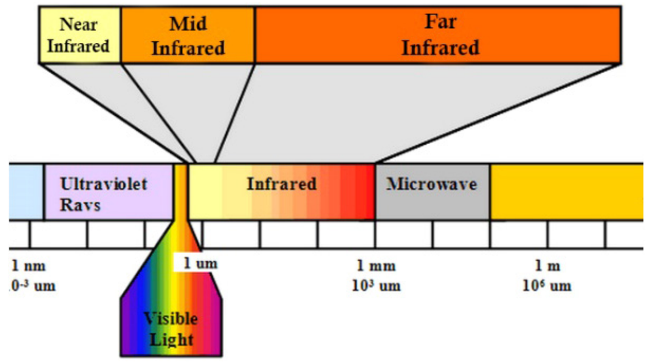


Fig. 3. Multispectral bands.

Fig. 4. The kinect sensor which was released by Microsoft.

2.3. *Kinect sensor*

The Kinect appeared on November 4, 2010, as an accessory to Xbox 360 console. It is a device developed by PrimeSense in collaboration with Microsoft. In January of 2012, more than 18 million units were sold. In February of 2012, a version for Windows was released [18,22].

The Kinect sensor is able to provide the RGB image, IR image and depth image simultaneously. Figure 5 shows different images taken from the Kinect sensor.

The sensor is used in Xbox 360 gaming console for control free game experience.

3. 3D Estimation Using Multispectral Vision

3D reconstruction from stereo images refers to the technique of obtaining a 3D representation of a scene from a set of images under different viewpoints. Furthermore, the problem of 3D estimation consists in combining several 2D representations of the same area to reconstruct it in three dimensions.

In our work, we propose three approaches for 3D estimation using different images acquired from the Kinect sensor under

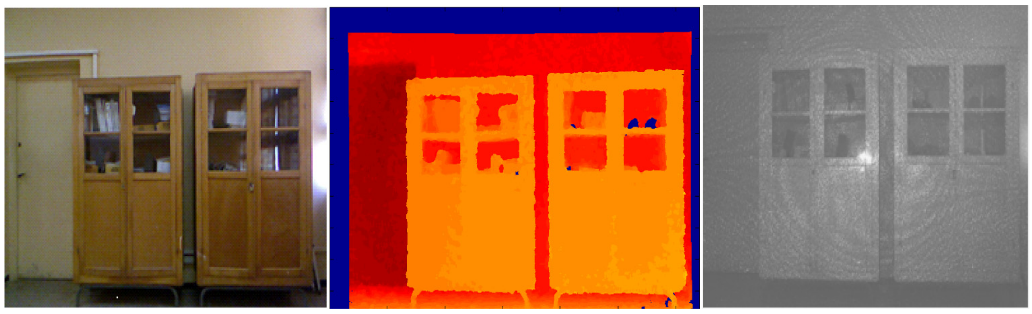


Fig. 5. Kinect provided data: RGB image (left), depth image (middle) and infrared image (right).

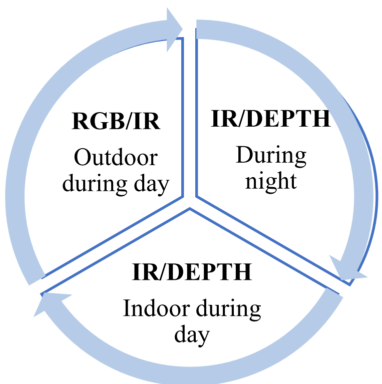


Fig. 6. Multimodal 3D estimation (day, night, indoor and outdoor).

different conditions (day, night, indoor and outdoor), according to Fig. 6.

As shown in Fig. 6, in indoor environment and during day, the combination of RGB and depth images is used for 3D estimation; this combination is allowed only when the depth image is available, and this is possible only in indoor environment because the IR infrared projector is highly parasitized by sunlight.

During nighttime as well as in dark area, the quality of the RGB image decreases significantly. In this case, the only solution for 3D estimation is to combine IR and depth images to estimate the 3D coordinate of the environment.

The major limitation of the Kinect sensor is its sensitivity to face sunlight, which makes it inefficient in outdoor environment. The presence of sunlight infrared radiation highly disturbs the Kinect IR projector leading to inaccurate depth measure. As a solution, we proposed a multispectral stereovision algorithm that combines RGB and IR images to estimate the features' depth. One of the major challenges in computer vision [23], and stereovision in particular, is the correspondence/matching problem [24]. This task is proved possible when applied on images of the same spectral band and very challenging on images captured with sensors operating in different spectrum such as visible (RGB) and infrared (IR) [25]. For this reason, we proposed to use features' extraction matching based on local invariant, and good results of features' correspondence are obtained, after a step of processing of the IR image. This step is used to improve the quality of the IR image, in order to facilitate features' extraction and matching with the RGB one, (Fig. 7 shows the previous steps). This treatment is mainly based on scale transformation and histogram adjustment.

To evaluate the average depth precision for ten (10) points located at the same distance from the Kinect sensor. To realize this experiment, we have fixed the Kinect on a support facing a checkerboard. We have moved the Kinect sensor along the Z axis and measured the real distance between the camera and the 10 points fixed on the checkerboard each time. Several tests have been carried out by

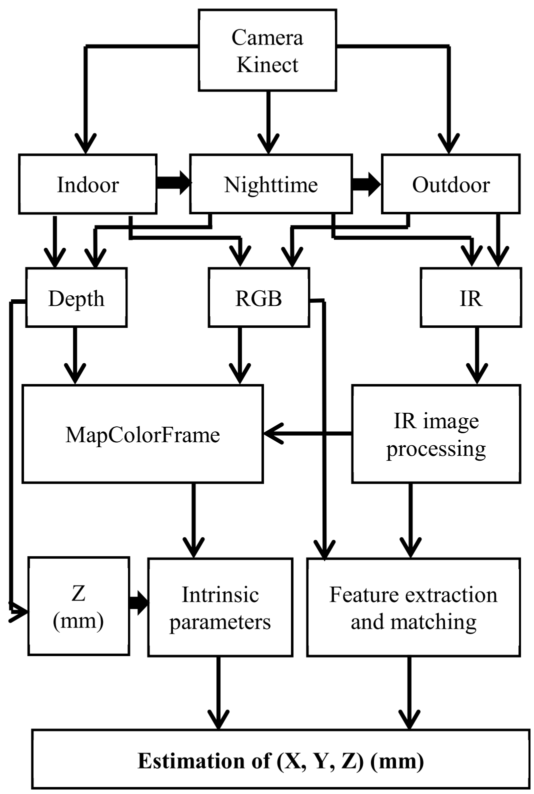


Fig. 7. 3D estimation using (RGB/depth/IR) images.

Table 1. 3D estimating accuracy.

Position (mm)	Day indoor RGB/DEPTH $Z \in [800\ 2000]$	Day outdoor RGB/IR $Z \in [1400\ 2200]$	Night IR/DEPTH $Z \in [800\ 2000]$
X (mm)	$\sigma_x = 07$	$\sigma_x = 15$	$\sigma_x = 10$
Y (mm)	$\sigma_y == 10$	$\sigma_y = 20$	$\sigma_y = 15$
Z (mm)	$\sigma_z = 30$	$\sigma_z = 37$	$\sigma_z = 33$

moving the sighting device following the Z axis from 800 mm to 2800 mm.

The following table summarizes the obtained results of 3D estimating accuracy in the three aforementioned cases. X, Y and Z as well as the standard deviation are estimated using 100 points with known coordinates.

Precision of estimation for the three approaches of depth estimation is illustrated in Table 1. Overall, 3D estimation in indoor, outdoor and during night is accurate and suitable for localization and navigation of unmanned vehicles. A depth error smaller than 37 mm is obtained for a distance of 2400 mm which corresponds to a relative error of 1.85% which represents a good precision for mobile robot navigation. In order to maintain a good localization, during navigation the observed points with depth > 2400 mm will not be considered. In the next section, visual odometry based on 3D point estimation is developed for mobile robot localization and navigation using the Kinect sensor.

4. Multispectral Visual Odometry

Visual odometry is defined as the process of estimating the robot's motion (translation and rotation with respect to a reference frame) by observing a sequence of images of its environment.

For every new image I_k (or image pair in the case of a stereo camera or Kinect), the first two steps consist of detecting and matching 2D features in consecutive frames. Image features that are the reprojection of the same 3D features across different frames are called 2D/3D image correspondences [26].

4.1. *Multispectral visual odometry for multimodal navigation*

In this paper, the localization algorithm based on MVO using the Kinect sensor is proposed for Multimodal Navigation (MMN) [27]. Three modes of navigation are considered: indoor navigation, outdoor navigation and nighttime navigation. For every new observation given by the Kinect, a 3D estimation (P_k^i) of the observed features is obtained using the available images at time i, according to the navigation mode (Fig. 6). The next step of the MVO is the estimation of the robot pose (translation and rotation) by minimizing the criterion through the following equation [28],

$$F(R,t) = \frac{1}{N}\sum_{k=1}^{N} \|(RP_k^{i+1} + t) - P_k^i\|^2. \tag{4}$$

4.2. *Multispectral visual odometry algorithm based on ICP-SURF*

To minimize the criterion given in Eq. (4), an iterative closest point (ICP) algorithm combined with a SURF detector/descriptor is proposed. In this case, the pose estimation consists in two steps. The first step involves finding the correspondence between two consecutive 3D scans using SURF detection and matching. The second step is to estimate the geometric transformation (rotation and translation) by minimizing the mean square error (MSE) between two consecutive scans (Fig. 8). Using the SURF descriptor for points' association significantly decreases the computation time of the ICP algorithm, and the estimated pose will satisfy the convergence criterion quickly.

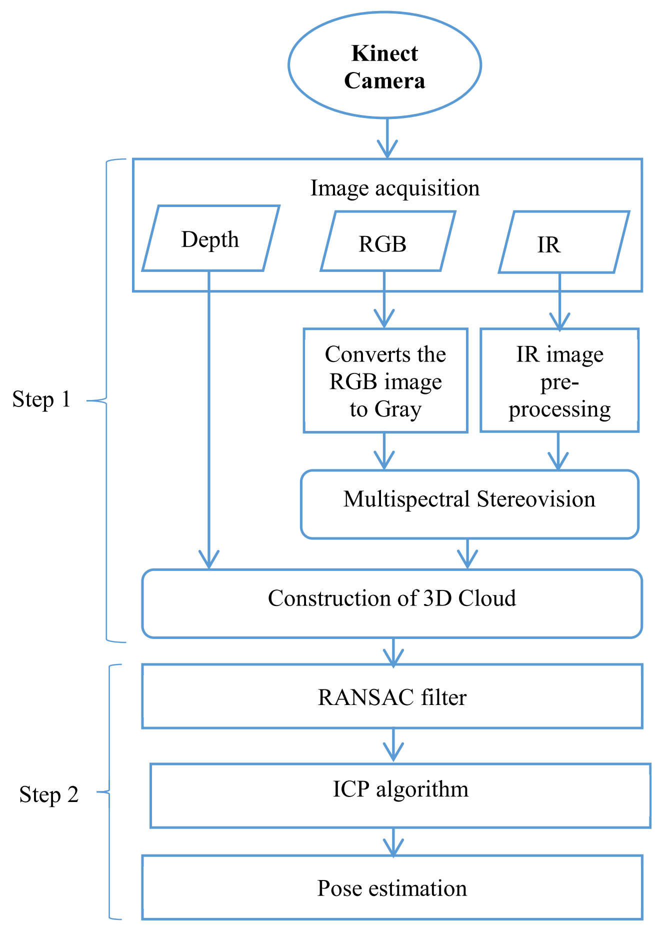


Fig. 8. MVO algorithm using ICP-SURF algorithm.

The principal of iterative closest SURF points' (ICP-SURF) algorithm for visual odometry [29,30] is represented in the following flowchart:

5. Data Fusion for Localization

In certain scenarios, in which the visual information is not available (bad weather or dark area), few number of SURF features are detected; thus the MVO algorithm [27] provides poor localization performances. Furthermore, when the mobile robot switches from the navigation mode to another navigation mode, the MVO may lose the estimated position. As a solution, in this paper, we propose to combine the MVO pose with the mobile robot WO pose using an extended Kalman filter [31,32]. The use of an additional sensor (WO) can maintain a suitable pose estimation and overcome MVO limitations. Moreover, the proposed EKF-MVO algorithm allows the localization algorithm to maintain an accurate position when switching between different localization modes (Fig. 9).

5.1. *Mobile robot localization using EKF-MVO*

The EKF is a well-known nonlinear version of the linear KF Algorithm [33]; it adopts the linearization method of the nonlinear function which takes the Taylor expansion to approximate the battery nonlinear function into a linear function [34,35],

$$X_K = f(X_{K-1}, U_{K-1}) + W_K, \tag{5}$$

$$Z_K = h(X_k) + V_K. \tag{6}$$

The model and implementation equations for extended Kalman filter are defined as follows:

Prediction:

$$\hat{X}_{K/K-1} = f(\hat{X}_K), \tag{7}$$

$$P_{K/K-1} = F(\hat{X}_K) P_K F^T(\hat{X}_K) + Q. \tag{8}$$

Update:

$$K_K = P_{K/K-1} H^T (H \cdot P_{K/K-1} \cdot H^T + R)^{-1}, \tag{9}$$

$$\hat{X}_K = \hat{X}_{K/K-1} + K_K[Z_K - h(\hat{X}_{K/K-1})], \tag{10}$$

$$P_K = P_{K/K-1} - K_K H P_{K/K-1}, \tag{11}$$

where $\hat{X}_k$ is the predicted state vector, $X_{k/k-1}$ is the estimated state vector, K is the Kalman gain matrix, P_k is the covariance matrix, F is the Jacobian matrix of the state equation f with respect to the state vector X_{k+1} and the noise variable W_k, H_{k+1} is the Jacobian matrix of the measurement h with respect to the state vector X_{k+1} and the noise variable V_k [34].

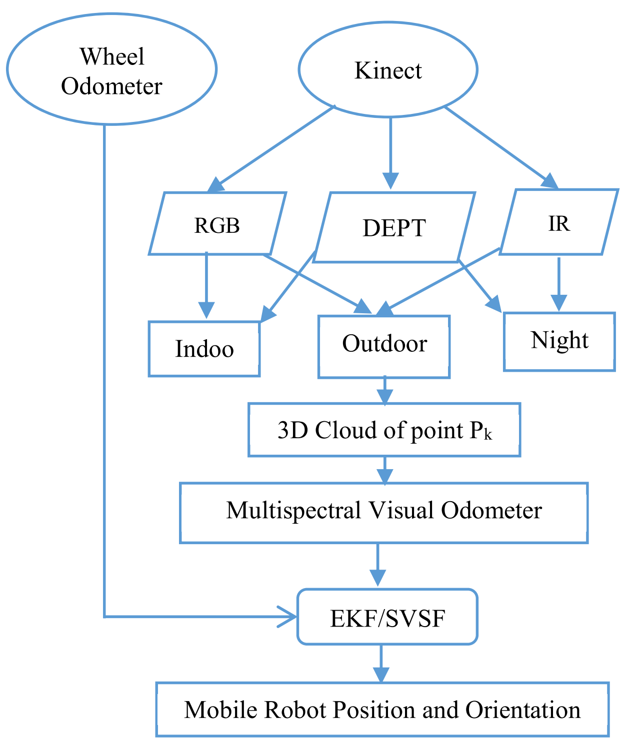


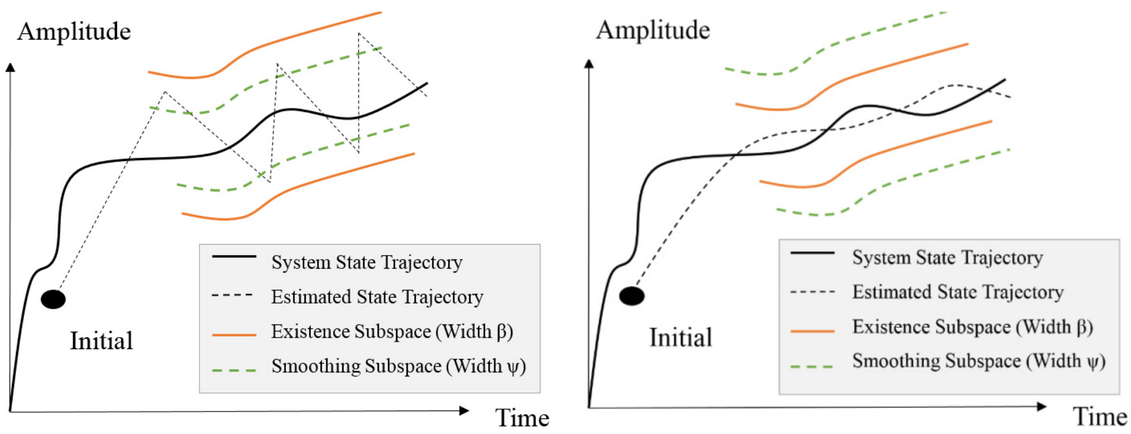
Fig. 9. Multimodal localization using multispectral visual odometer.

5.2. *Mobile robot localization using SVSF-MVO*

The SVSF is a predictor-corrector filter. This filter is based on the sliding mode control and estimation techniques, mostly used for the state and parameter estimation of a dynamic system whether linear or nonlinear, accurate fault detection, estimation and tracking applications.

The SVSF defines a hyperactive plane, which is referred to as the sliding surface and applies a discontinuous force on the estimate to make it reach the hyperactive plane in two stages: prediction and update stages. The estimates in each stage remain within the subspace surrounding the sliding hyperactive plane, referred to as the existence subspace as illustrated in Fig. 10. During each stage, an artificial noise referred to as chattering is created depending on uncertainties and the SVSF coefficient matrix Υ. Chattering obtained in the prediction stage is referred to as the secondary indicator of performance. Conversely, chattering in the a posteriori estimate decays with time by the factor Υ^k; therefore no information could be obtained.

The goal of SVSF is the attitude estimation of the mobile system by fusing inertial data with the three GPS antenna data. The convergence of the SVSF results in a switching around the system state. The way of switching depends on the smoothing subspace presented by the width δ, and the existence subspace presented by the width β, which represents the model uncertainties (modeling errors and



(a) (b)

Fig. 10. The SVSF concepts.

measurement noise). If $\delta \geq \beta$, the estimated state trajectory is smoothed (Fig. 3). Whereas, if $\delta < \beta$, the chattering phenomenon appears owing to the underestimation of uncertainties (Fig. 4).

The estimation process is iterative and may be summarized by the following set of equations [36,37]:

The predicted state:

$$\hat{X}_{k+1/k} = f(\hat{X}_{k/k}, U_k), \tag{12}$$

the corresponding predicted measurement:

$$\hat{Z}_{k+1/k} = h(\hat{X}_{k+1/k}), \tag{13}$$

the *a priori* measurement error:

$$E_{k+1/k} = Z_{k+1} - \hat{Z}_{k+1/k} \tag{14}$$

and the (SVSF) gain:

$$\begin{aligned} K_{k+1} = {} & (H_{k+1})^{+}\mathrm{diag}[(|E_{k+1/k}| + \gamma|E_{k/k}|)^{\circ} \\ & \times sate(\bar{\varphi}^{-1}E_{k+1/k})] * [\mathrm{diag}(E_{k+1/k})]^{-1}, \end{aligned} \tag{15}$$

where

H: the linearized measurement matrix,

γ: the convergence rate matrix with $((0 \leq \gamma_{ii} \leq 1))$,

φ: the smoothing boundary layer width,

$^{\circ}$: the Schur (or element-by-element) multiplication, the superscript,

$+$: the pseudoinverse of a matrix and

$\bar{\varphi}^{-1}$: the diagonal matrix constructed from the smoothing boundary layer vector φ,

with

$$sate(\bar{\varphi}^{-1} * E_{k+1/k}) = \begin{cases} 1 & \varphi_i^{-1} * E_{k+1/k} \geq 1 \\ \bar{\varphi}_i^{-1} E_{k+1/k} & -1 < \varphi_i^{-1} * E_{k+1/k} < 1\,. \\ -1 & \varphi_i^{-1} * E_{k+1/k} \leq -1 \end{cases} \tag{16}$$

The updated state estimate:

$$\hat{X}_{k+1/k+1} = \hat{X}_{k+1/K} + K_{k+1} * E_{k+1/k}. \tag{17}$$

The updated measurement estimate:

$$Z_{k+1/k+1} = h(\hat{X}_{k+1/k+1}). \tag{18}$$

The posterior measurement error:

$$E_{k+1/k+1} = Z_{k+1} - \hat{Z}_{k+1/k+1}. \tag{19}$$

6. Automatic Selection of Navigation Mode

The main challenge in multimodal navigation (MMN) is the switching condition between different modes (indoor, outdoor and nighttime). The mobile robot should automatically detect which mode of navigation is required following the availability of data. In our work image, energy was used as criteria to automatically switch between localization modes [38]. For example, during nighttime, the energy of the RGB image decreases significantly less than the threshold; thus MVO-SVSF and MVO-EKF algorithms based on RGB/depth and RGB/IR cannot be used; on the other hand, the MVO-EKF and MVO-SVSF based on IR/depth images are more appropriate in this navigation mode.

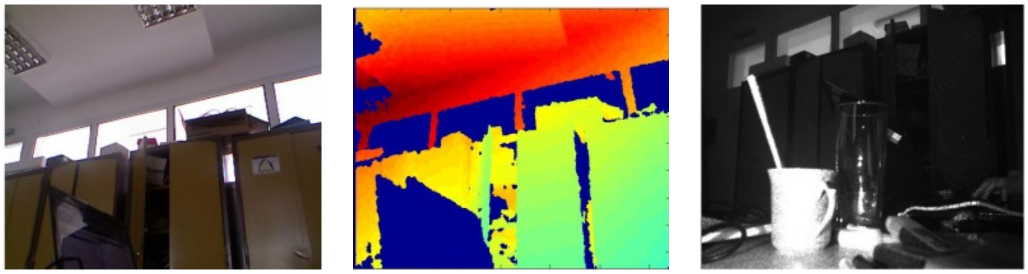
6.1. *Images energy calculation*

In this paper, image energy is used as criteria to automatically switch between localization modes. The energy of image I of size $N \times M$ can be calculated as follows:

$$\begin{aligned} \text{Image energy} &= E \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{j} I(i,j)^2/(255 * \mathrm{N} * \mathrm{M}). \end{aligned} \tag{20}$$

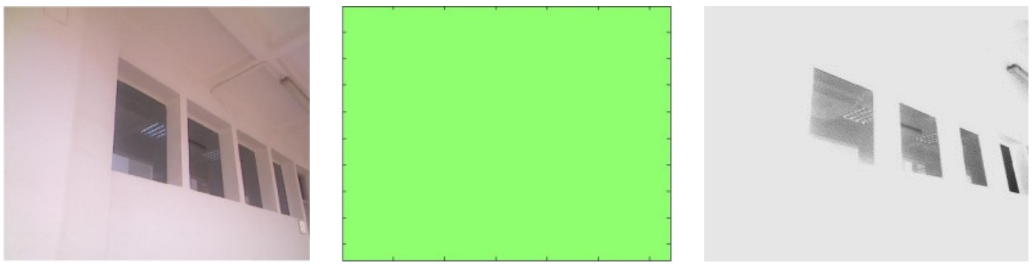
In this case, the image energy is normalized between $[0, 1]$.

June 7, 2022 12:01:28pm WSPC/284-US 2250015 ISSN: 2301-3850 2nd Reading



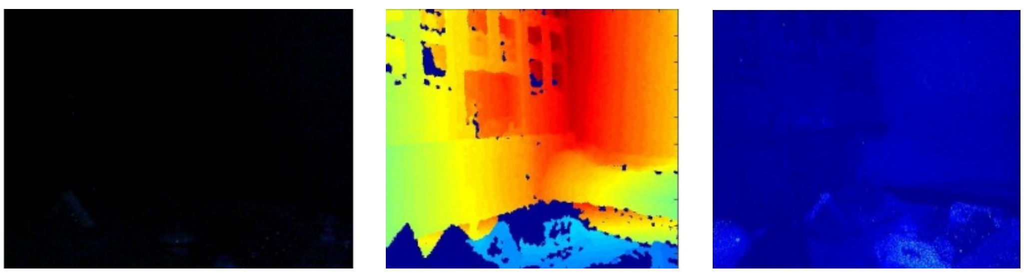
(a) $E_{RGB} = 0.9588$ (b) $E_{depth} = 0.7363$ (c) $E_{IR} = 0.1385$

Fig. 11. Indoor localization: (a) RGB image, (b) IR image and (c) depth image.



(a) $E_{RGB} = 0.5982$ (b) $E_{depth} = 00$ (c) $E_{IR} = 0.7974$

Fig. 12. Outdoor localization: (a) RGB image, (b) IR image and (c) depth image.



(a) $E_{RGB} = 0.0194$ (b) $E_{depth} = 0.8831$ (c) $E_{IR} = 0.1179$

Fig. 13. Night localization: (a) RGB image, (b) IR image and (c) depth image.

Figure 11 illustrates three images (RGB, depth and IR) acquired by the Kinect sensor in indoor environment; as can be seen, RGB and depth images have significant energy compared to the IR image energy. In this navigation mode, the EKF-MVO based on RGB/depth is more suitable for mobile robot localization. Images acquired (RGB, depth and IR) in the outdoor navigation mode are presented in Figs. 12(a)–12(c). As can be seen, the RGB and IR images have significant energy compared to the depth image energy (the IR projector is highly affected by the IR sunlight). In this navigation mode, the EKF-MVO based on RGB/IR is more suitable for mobile robot localization. During nighttime, as shown in Fig. 13, depth and IR images have significant energy compared to the RGB image energy. In this navigation mode, the EKF-MVO based on depth/IR is more appropriate for mobile robot localization.

When visual information is not available, the MVO localization is not possible and the WO will be used to localize the mobile robot.

Table 2 illustrates the suitable navigation mode following image energies (E_{RGB}: energy of the RGB image, E_{Depth}: energy of the depth image, and E_{IR}: energy of the IR image).

6.2. *Automatic navigation mode selection using fuzzy logic controller*

To ensure automatic and smooth switching between navigation modes when a new environment is detected, a switching FLC is proposed. The switching FLC has three inputs (images' energies) and one output (navigation mode) as illustrated in (Fig. 14), where M1–M4 represent the navigation mode as illustrated in Table 3.

Table 2. Relation between image energy and visual navigation mode.

Energy	$E_{\mathrm{RGB}} < \mathrm{th}_{\mathrm{RGB}}$	$E_{\mathrm{depth}} < \mathrm{th}_{\mathrm{depth}}$	$E_{\mathrm{IR}} < \mathrm{th}_{\mathrm{IR}}$	$E_{\mathrm{RGB}} < \mathrm{th}_{\mathrm{rgb}}$ & $E_{\mathrm{depth}} < \mathrm{th}_{\mathrm{depth}}$ & $E_{\mathrm{IR}} < \mathrm{th}_{\mathrm{IR}}$
Algorithms	MVO-Filter depth/IR	MVO-Filter RGB/IR	MVO-FilterRGB/Depth WO	WO
Navigation modes	M1	M2	M3	M4

Notes: $\mathrm{Th}_{\mathrm{RGB}}$, $\mathrm{th}_{\mathrm{depth}}$ and $\mathrm{th}_{\mathrm{IR}}$ are the energy threshold of RGB, depth and IR image, respectively. M1, M2, M3 and M4 represent the navigation modes based on the available visual data.

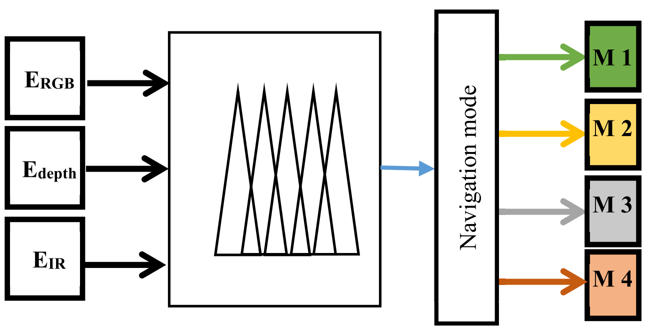


Fig. 14. Switching FLC block for navigation mode selection.

6.2.1. *FLC inputs*

The membership functions of the FLC inputs (for RGB, depth and IR energies) are illustrated in Figs. 15(a)–15(c).

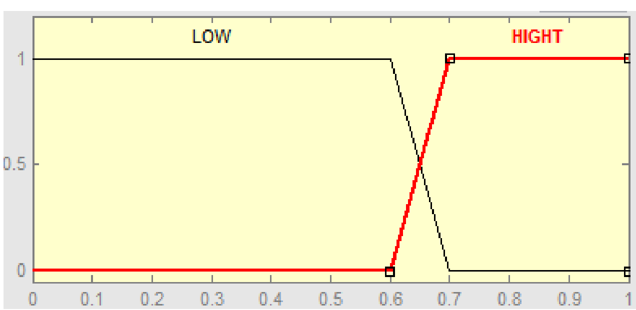
The controller output is illustrated in Fig. 16, where four navigation modes can be selected as shown in Table 3. The mode M4 is the mobile robot WO; this mode is activated when visual information is not available.

6.2.2. *Fuzzy rules*

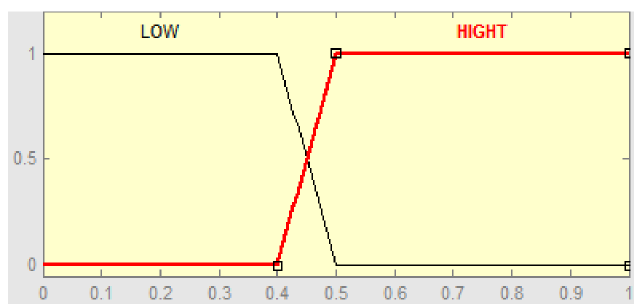
The inference matrix (fuzzy rules) of the proposed controller is illustrated in Table 3. This table is developed based on eight (08) IF-THEN fuzzy rules.

7. Results and Discussion

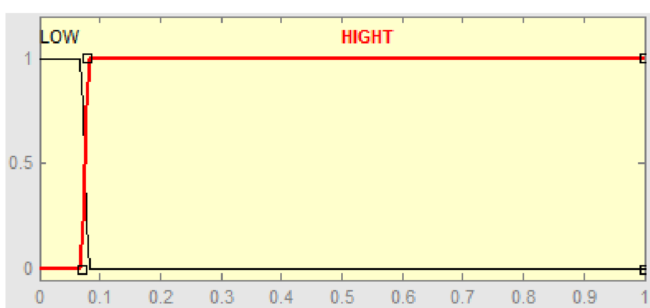
In Sec. 3, 3D estimation algorithms are validated and evaluated in different navigation modes. Based on these results, further experiments are considered to evaluate and validate the proposed MVO algorithm for mobile robot localization. The localization algorithm is validated on a trajectory-tracking task, where the mobile robot with an embedded Kinect sensor (Fig. 17) should accurately track a desired (reference) trajectory using the proposed approaches.



(a) Input variable "E_{RGB}"



(b) Input variable "E_{DEPTH}"



(c) Input variable "E_{IR}"

Fig. 15. Membership function of FLC inputs.

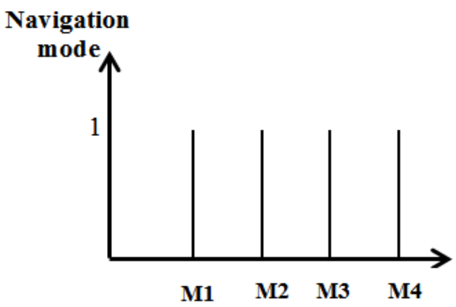


Fig. 16. Controller output of FLC automatic switching.

Table 3. Fuzzy inference matrix for localization mode selection.

E_{RGB}	E_{IR}	E_{Depth}	Output
Low	Low	Low	Mode 4: Odometer
Low	Low	High	Mode 4: Odometer
Low	High	Low	Mode 4: Odometer
Low	High	High	Mode 3: Night (IR/Depth)
High	Low	Low	Mode 4: Odometer
High	Low	High	Mode 1: Indoor (RGB/Depth)
High	High	Low	Mode 2: Outdoor (RGB/IR)
High	High	High	Mode 1: Indoor (RGB/Depth)

Fig. 17. The mobile pioneer robot with a mounted Microsoft Kinect camera.

Different navigation modes for the mobile robot are considered in indoor, outdoor and during night time. The proposed EKF-MVO is compared to the MVO and the mobile robot WO.

In this experiment, we consider a trajectory tracking for multimodal navigation (MMN). In this case, the mobile robot navigates through different environments under different conditions and several localization modes are required for mobile robot localization and suitable trajectory tracking (Fig. 18).

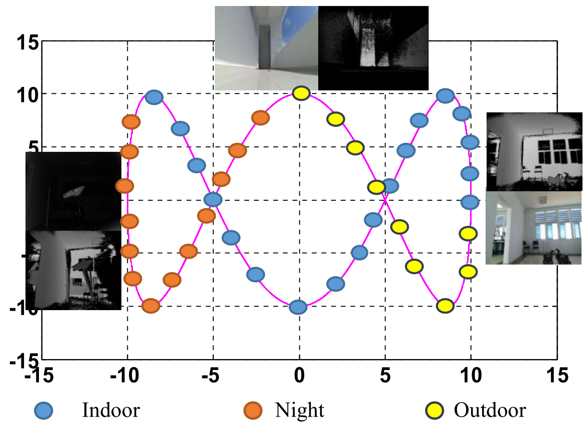


Fig. 18. Multimodal localization and navigation.

First, when $T \in [0, 400\,\mathrm{s}]$ the mobile robot is in indoor, then, when $T \in [400\,\mathrm{s}, 800\,\mathrm{s}]$ the mobile robot explores an outdoor environment, and finally, for $T \in [800\,\mathrm{s}, 1250\,\mathrm{s}]$ the mobile robot crosses a dark area. Trajectory tracking results using different localization modes are illustrated blow.

7.1. *Multimodal localization and navigation based on MVO-EKF*

Trajectory tracking results using different localization modes are illustrated in Fig. 19. As can be seen from this figure, the MVO-EKF provides good tracking precision comparing to the MVO localization system. The FLC system ensures a smooth switching between different navigation modes.

Mobile robot poses (X, Y and Theta) are illustrated in Fig. 20. Figures 21(a)–21(c) show that the MVO-EKF

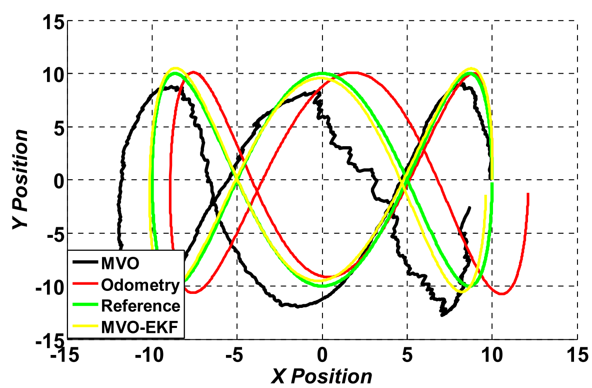
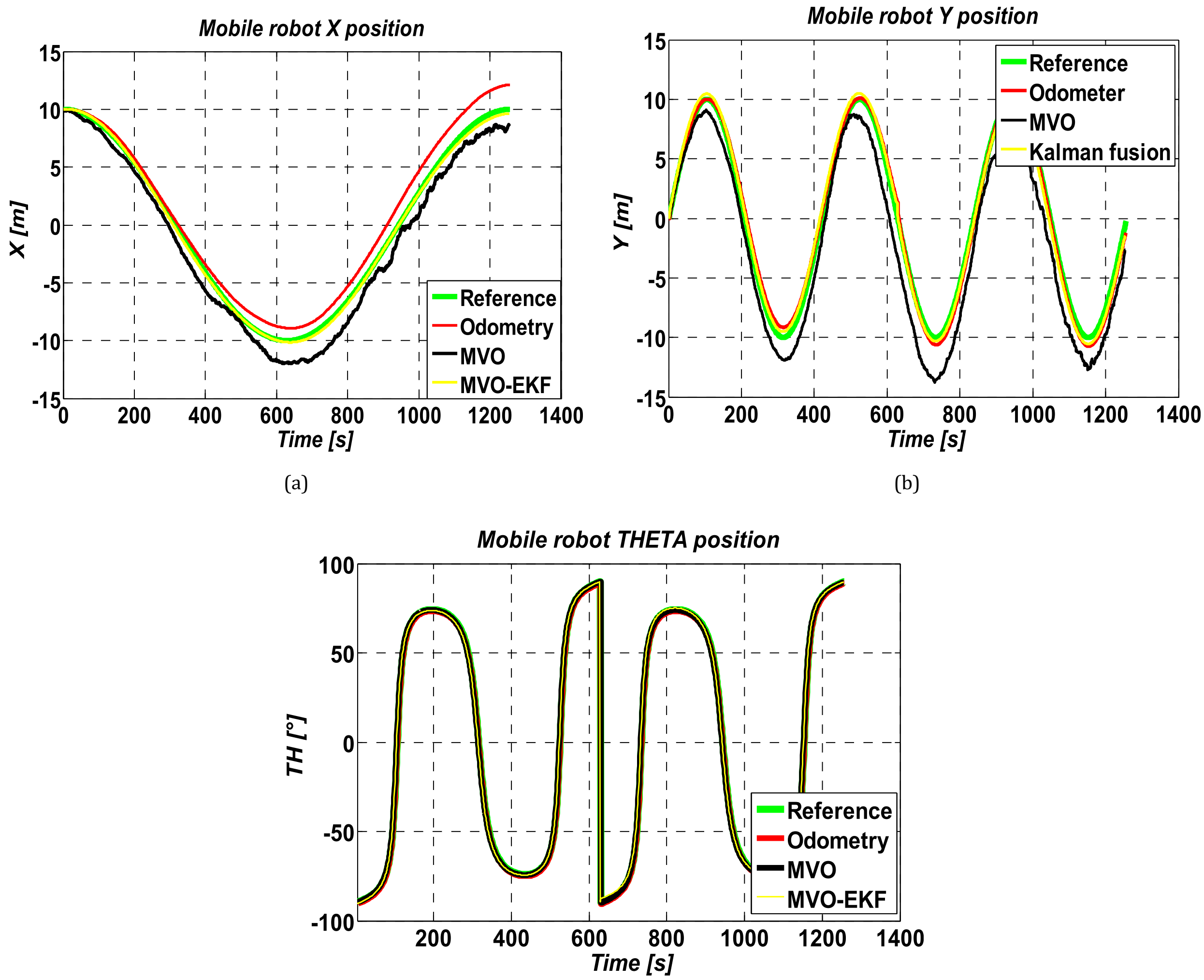


Fig. 19. Trajectory tracking using MVO, EKF-MVO and odometry for long-term localization (multimodal localization).



(a) (b) (c)

Fig. 20. (a) Mobile robot X position, (b) mobile robot Y position, (c) mobile robot THETA position.

algorithm provides an accurate mobile robot pose when the odometry diverge and the MVO diverge gradually during the navigation mode change. The tracking algorithm based on MVO-EKF localization maintains a good tracking precision; when the navigation environment changes, the switching FLC maintains accurate and smooth pose estimation, even when the navigation mode changes.

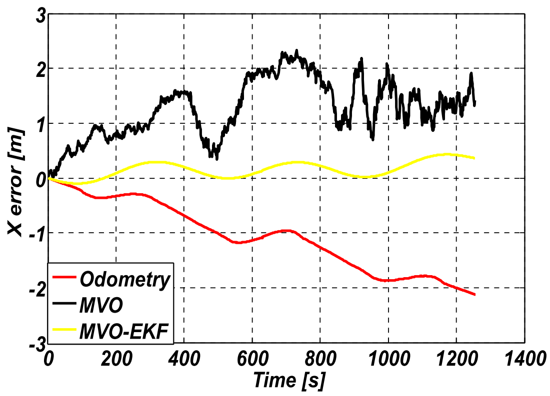
7.2. *Multimodal localization and navigation based on MVO-SVSF*

***Experiment* 1**

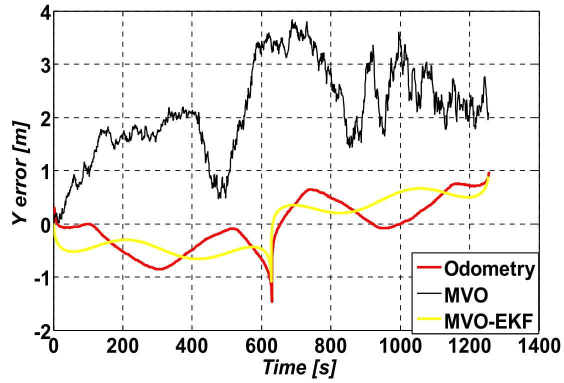
Figure 22 represents the obtained results' indifferent localization modes using the three localization approaches (robot odometry, MVO and MVO-SVSF) that are implemented and compared to the desired trajectory. As can be seen from this figure, the MVO-SVSF localization is more accurate and smooth compared to the MVO localization, when the mobile robot odometry position drifts significantly with time.

Figures 23(a)–23(c) show the mobile robot position and orientation (X, Y and Theta) for different localization systems; precision evaluation is given in Figs. 24(a)–24(c), where the MVO-SVSF's error following X, Y and *Theta* is smaller compared to MVO and odometry errors. The latter drifts significantly with time.

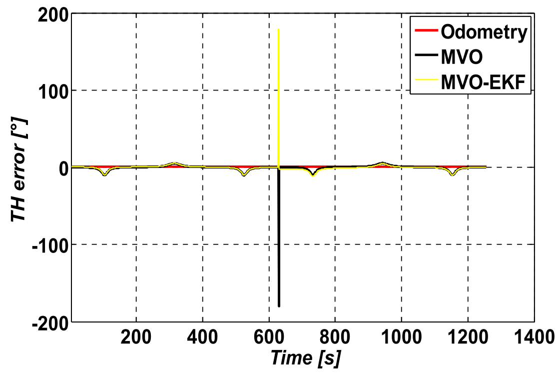
Table 4 illustrates a comparative study (precision/computational time) between the MVO, MVO-EKF and MVO-SVSF. As can be seen, the best performances have been obtained with the proposed MVO-SVSF. Moreover, the MVO-SVSF performs much better than the MVO-EKF, which requires a linearization step for pose estimation. The linearization process increases the computational time and makes the EKF less precise than the SVSF.



(a)



(b)



(c)

Fig. 21. (a) X error, (b) Y error and (c) THETA error.

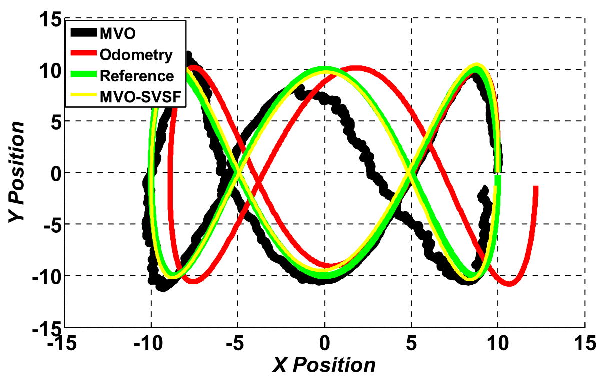
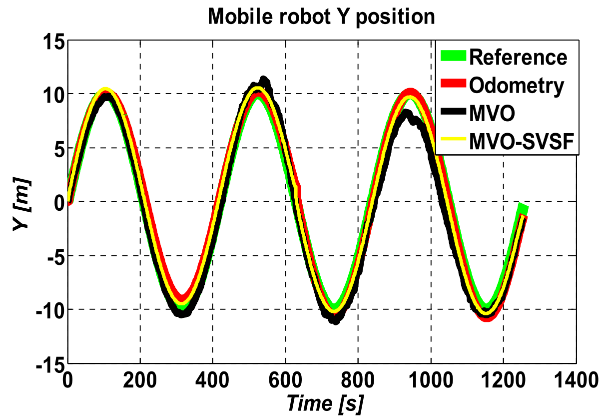


Fig. 22. Trajectory tracking using MVO, MVO-SVSF and odometry for long-term localization (multimodal localization).

Experiment 2

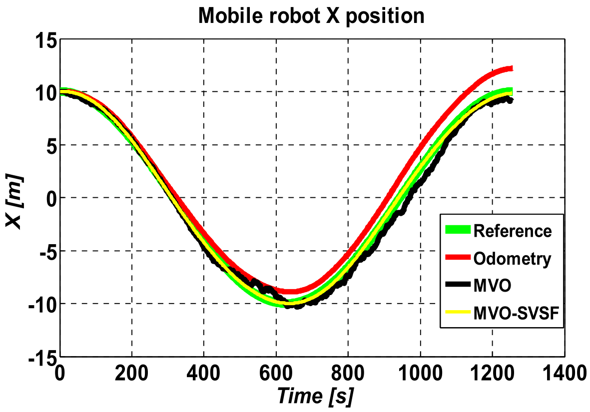
In this experiment, the mobile robot P3-AT will navigate in a large area of the LVAI laboratory (Fig. 25); the blue points represent the indoor area while the yellow points represent the outdoor area.

Figure 26 represents the obtained results in different localization modes using the two localization approaches (robot odometry and MVO) that are implemented and compared to the desired trajectory. As can be seen from this figure, the MVO localization is more accurate and smooth compared to the WO localization, when the mobile robot odometry position drifts significantly with time.

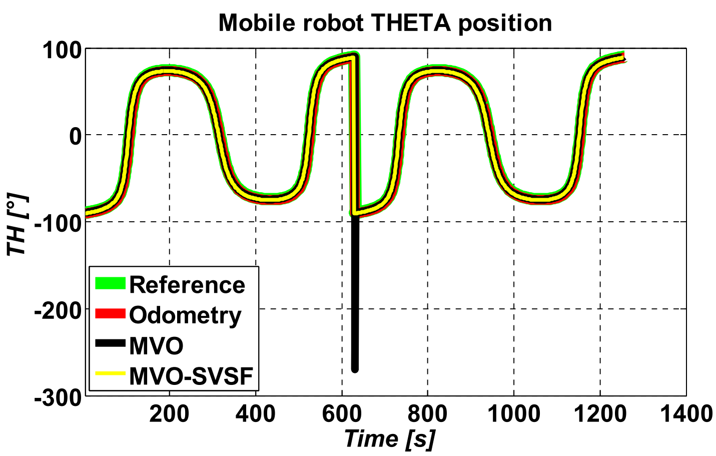


(a)

Fig. 23. (a) Mobile robot X position, (b) mobile robot Y position and (c) mobile robot Theta position.



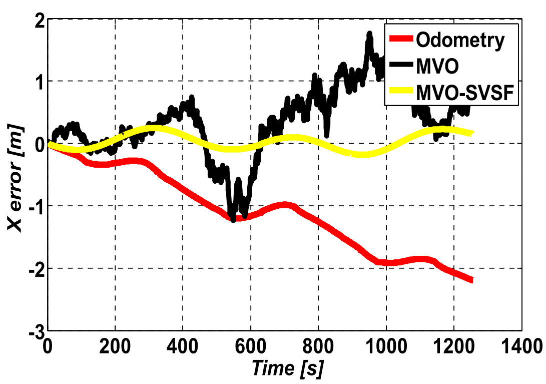
(b)



(c)

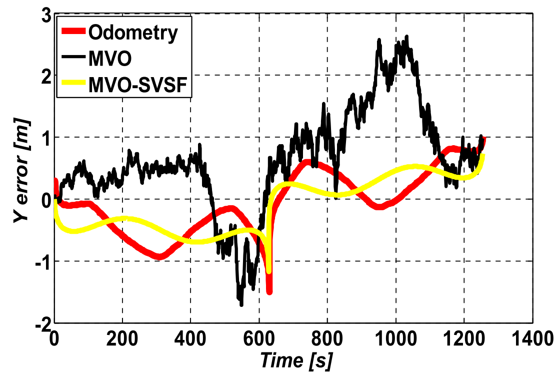
Fig. 23. (*Continued*)

Table 5 summarizes some work already done in each specific environment compared to our approach. As we see in Ref. 39, the authors propose an approach for outdoor localization, based on GPS and stereovision between visible

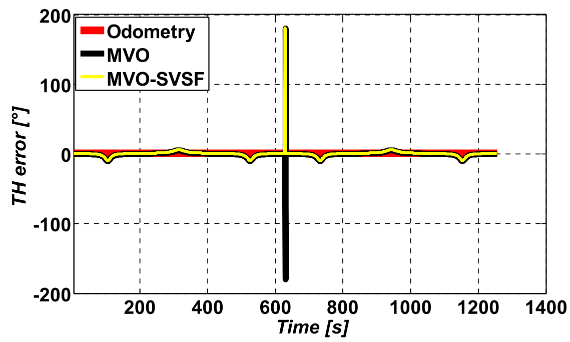


(a)

Fig. 24. (a) X error, (b) Y error and (c) Theta error.



(b)



(c)

Fig. 24. (*Continued*)

Table 4. Consumed time and RMSE of odometry, MVO, MVO-EKF and MVO-SVSF.

	Odometry	MVO	MVO-EKF	MVO-SVSF
X-RMSE (m)	1.08	0.60	0.095	0.082
Y-RMSE (m)	0.60	0.80	0.30	0.25
θ-RMSE (°)	0.90	0.60	0.80	0.80
Frame per second (fps)	/	13.72	14.79	15.41
Sampling time (ms)	28.54	72.86	67.6	64.86

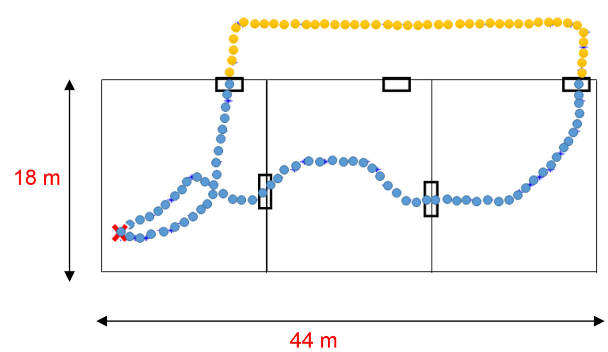


Fig. 25. Laboratory area where the experiment was done (blue: indoor, yellow: outdoor).

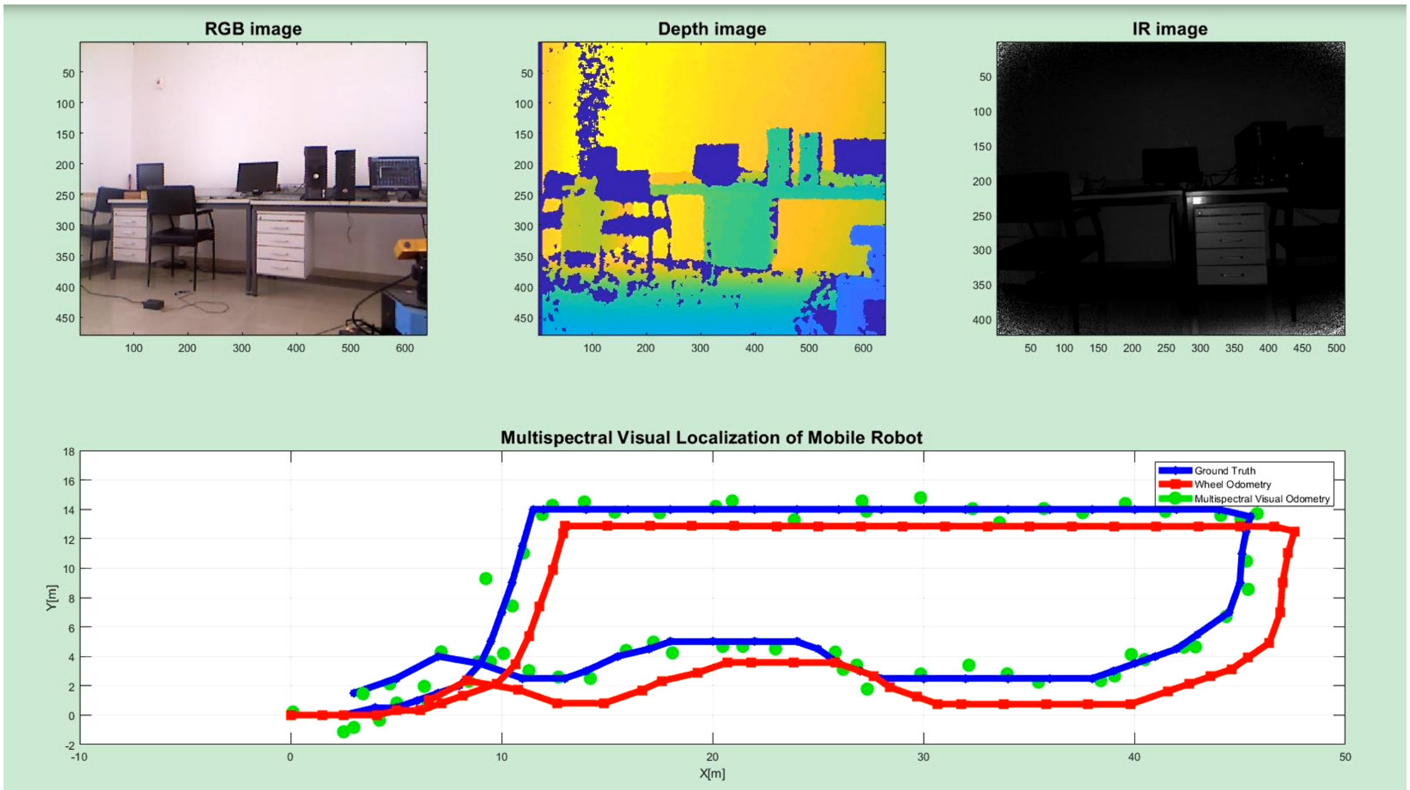


Fig. 26. Top: multispectral images, bottom: MVO localization and navigation in indoor and outdoor.

Table 5. Comparison of our experiment with previous methods in each specific environment.

Environments	Indoor	Outdoor	Night localization	Accuracy (%)
Axel Beauvisage *et al.*[39]	−	+	−	6.87
Tarek Mouats *et al.*[40]	−	+	−	8.50
CHAIB Khaoula[20]	+	−	−	6.33
Asha Anoosheh *et al.*[41]	−	−	+	9.10
Our approach	+	+	+	3.30

and Long Wave InfraRed (LWIR) spectral bands. Another approach for the same environment is proposed by TarekMouats *et al.* in Ref. 40, where a stereovision algorithm based on optical (visible) and thermal cameras is implemented to localize an autonomous vehicle.

In Ref. 20, the authors combined the RGB and IR images acquired from the Kinect sensor for indoor localization. Images of daytime and nighttime are fused by AshaAnoosheh et Al in Ref. 41, to ensure localization at nighttime.

We can see that our approach ensures the localization in different conditions and it yields a better estimate of the trajectory. In accuracy (%), our approach is about 3.30, whereas the other approaches are about 6.33–9.10.

Another solution has been proposed in Ref. 13 which consists in computer stereovision for autonomous driving, where the authors talk about visual feature detection, description and matching. Xuyou Li *et al.* in Ref. 13, proposed to use 6DoF simultaneous localization and mapping (SLAM) using the standard ICP method and its variants. A new algorithm is illustrated in Ref. 14, where the authors propose a multimodel EKF integration navigation algorithm and estimate the measurement errors of beacons online.

8. Conclusion

In this paper, an adaptive algorithm of multimodal localization using MVO for mobile robot navigation is proposed. The first objective was to implement a robust algorithm for 3D estimation using the Kinect sensor in different conditions: day, night, indoor and outdoor. A multimodal localization system using a fuzzy adaptive MVO is proposed for switching between different navigation modes. To overcome the limitation of the multimodal MVO especially during the navigation mode switching, a SVSF and an EKF are implemented and compared with MVO/WO data fusion.

The proposed localization algorithms are evaluated and validated on a trajectory tracking problem. From the experimental results, good precision of 3D estimation has been obtained by the proposed approach even in outdoor environment. The best multimodal localization performances for indoor, outdoor, and nighttime localization are obtained using fuzzy adaptive MVO-SVSF and MVO-EKF, with superiority for MVO-SVSF. The latter is able to maintain a suitable precision during navigation modes' switching without any assumption about noise characteristics.

References

[1] N. Morette, "Contribution a la navigation de robots mobiles: Approche par modèle direct et commande prédictive," Thèse Doctorat, Université d'Orléans (2009).

[2] Z.-H. Wang, K.-Y. Qin, T. Zhang and B. Zhu, An intelligent ground-air cooperative navigation framework based on visual-aided method in indoor environments, *Unmanned Syst.* **9** (2021) 237–246.

[3] S. A. S. Mohamed, M. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen and J. Plosila, A survey on odometry for autonomous navigation systems, *IEEE Access* **7** (2019) 97466–97486.

[4] T. Ort, L. Paull and D. Rus, Autonomous vehicle navigation in rural environments without detailed prior maps, *2018 IEEE Int. Conf. Robotics and Automation (ICRA)*, Brisbane, QLD, Australia, 2018, pp. 2040–2047, doi: 10.1109/ICRA.2018.8460519.

[5] M. O. A. Aqel *et al.*, Review of visual odometry: Types, approaches, challenges, and applications, *SpringerPlus* **5** (2016) 1897, https://doi.org/10.1186/s40064-016-3573-7.

[6] N. Ganganath and H. Leung, Mobile robot localization using odometry and kinect sensor, *2012 IEEE Int. Conf. Emerging Signal Processing Applications*, Las Vegas, NV, 2012, pp. 91–94, doi: 10.1109/ESPA.2012.6152453.

[7] H. Zhan, C. S. Weerasekera, J.-W. Bian and I. Reid, Visual Odometry Revisited: What Should Be Learnt?, *2020 IEEE Int. Conf. Robotics and Automation (ICRA)*, Paris, France, 2020, pp. 4203–4210, doi: 10.1109/ICRA40945.2020.9197374.

[8] A. S. Huang *et al.*, Visual odometry and mapping for autonomous flight using an RGB-D camera, in *Robotics Research*, eds. H. Christensen and O. Khatib (eds) Springer Tracts in Advanced Robotics, Vol. 100. (Springer, Cham, 2017), doi.org/10.1007/978-3-319-29363-9_14.

[9] A. Kadambi, A. Bhandari and R. Raskar, 3d depth cameras in vision: Benefits and limitations of the hardware, in *Computer Vision and Machine Learning with RGB-D Sensors*, (Springer, 2014), pp. 3–26.

[10] A. Ufkes and M. Fiala, Visual odometry using 3-dimensional video input, presented at the *Canadian Conf. Computer and Robot Vision*, Canada, 2011, pp. 86–93, doi:10.1109/CRV.2011.19.

[11] A. Alapetite, Z. Wang, J. P. Hansen, M. Zajączkowski and M. Patalan, Comparison of three off-the-shelf visual odometry systems, *Robotics* **9** (2020) 56, https://doi.org/10.3390/robotics9030056.

[12] X. Li, S. Du, G. Li and H. Li, Integrate point-cloud segmentation with 3D LiDAR scan-matching for mobile robot localization and mapping, *Sensors* **20** (2020) 237, https://doi.org/10.3390/s20010237.

[13] R. Fan, L. Wang, M. J. Bocus and I. Pitas, Computer stereo vision for autonomous driving, preprint (2020), arXiv:2012.03194.

[14] D. Li, D. Ji, J. Liu and Y. Lin, A multimodel EKF integrated navigation algorithm for deep water AUV, *Int. J. Adv. Robot. Syst.* **13**(1) (2016) 1, doi: 13.1.10.5772/62076.

[15] M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* **24** (1981) 381–395, doi.org/10.1145/358669.358692.

[16] P. Henry, M. Krainin, E. Herbst, X. Ren and D. Fox, RGB-D Mapping: Using depth cameras for dense 3D modeling of indoor environments, *Symp. Experimental Robotics*, New Delhi and Agra, India, December 2010, p. 79.

[17] F. Alhwarin, A. Ferrein and I. Scholl, IR stereo kinect: Improving depth images by combining structured light with IR stereo. In: D. N. Pham and S. B. Park (eds), *PRICAI 2014: Trends in Artificial Intelligence. PRICAI 2014*. Lecture Notes in Computer Science, Vol 8862. Springer, Cham, 2014, pp. 409–421, https://doi.org/10.1007/978-3-319-13560-1_33.

[18] D. Catuhe, *Programming with the Kinect for Windows Software Development Kit* (Microsoft Press, 2012).

[19] A. Abdo, R. Ibrahim and N. Rawashdeh, Mobile robot localization evaluations with visual odometry in varying environments using Festo-Robotino, SAE Technical Paper 2020-01-1022 (2020), https://doi.org/10.4271/2020-01-1022.

[20] C. Khaoula, Localisation et cartographie 3D en utilisant une camera RGB-D (Kinect), Magister, Ecole Militaire Polytechnique (2014).

[21] Y. T. Solano-Correa, F. Bovolo and L. Bruzzone, An approach for unsupervised change detection in multitemporal VHR images acquired by different multispectral sensors, *Remote Sens.* **10** (2018) 533, doi.org/10.3390/rs10040533.

[22] L. Cruz, D. Lucio and L. Velho, Kinect and RGBD Images: Challenges and applications, in *2012 25th SIBGRAPI Conf. Graphics, Patterns and Images Tutorials*, OuroPreto, 2012, pp. 36–49, doi: 10.1109/SIBGRAPI-T.2012.13.

[23] Z. Zhou, FPGA implementation of computer vision algorithm, A thesis submitted in partial satisfaction of the equirement for the degree of Master of Science in Electrical Engineering, University of California, Riverside (2014).

[24] A. Bhandari and K. C. Tiwari, Loss of target information in full pixel and subpixel target detection in hyperspectral data with and without dimensionality reduction, *Evolving Syst.* **12**(6) (2019) 239–254, doi.org/10.1007/s12530-019-09265-w.

[25] H. Kavitha and M. V. Sudhamani, Experimental analysis of SIFT and SURF features for multi-object image retrieval, *Int. J. Comput. Vis. Robot.* **7**(3) (2017) 344–356, doi: 10.1504/IJCVR.2017.083448.

[26] D. Scaramuzza and F. Fraundorfer, Visual odometry [Tutorial], *IEEE Robot. Autom. Mag.* 18(4) (2011) 80–92, doi: 10.1109/MRA.2011.943233.

[27] M. Grudziński, Ł. Marchewka, M. Pajor and R. Ziętek, Stereovision tracking system for monitoring loader crane tip position, *IEEE Access*, **8** (2020) 223346–223358, doi: 10.1109/ACCESS.2020.3043414.

[28] G. Caron, Estimation de pose et asservissement de robot par vision omnidirectionnelle, thèse de doctorat en science, UNIVERSITÉ DE PICARDIE JULES VERNE ÉCOLE DOCTORALE SCIENCES ET SANTÉ (France, 2010).

[29] T. H. Nguyen, T.-M. Nguyen, M. Cao and L. Xie, Loosely-coupled ultra-wideband-aided scale correction for monocular visual odometry, *Unmanned Syst.* **8** (2020) 179–190.

[30] A. Salleh, D. N. S. Dharmiza and S. Emmanuel, Swift path planning: Vehicle localization by visual odometry trajectory tracking and mapping, *Unmanned Syst.* **6** (2018) 221–230.

[31] T. A. Wenzel, K. J. Burnham and M. V. Blundell, Simplified extended Kalman filter for automotive state estimation, *Int. J. Model. Identification Control* **3**(3) (2008) 201–211, doi: 10.1504/IJMIC.2008.020120.

[32] L. A. Páramo-Carranza, J. A. Meda-Campaña, J. de JesúsRubio *et al.*, Discrete-time Kalman filter for Takagi–Sugeno fuzzy models, *Evolving Syst.* **8** (2017) 211–219, doi: 10.1007/s12530-017-9181-0.

[33] H. Zhao and Z. Wang, Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended Kalman filter for data fusion, *IEEE Sens. J.* **12**(5) (2012) 943–953.

[34] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox and N. Roy, Visual odometry and mapping for autonomous flight

using an RGB-D camera, *Robot. Res.* (Springer, Cham, 2017), pp. 235–252.

[35] Z. Zhang, H. Rebecq, C. Forster and D. Scaramuzza, Benefit of large field-of-view cameras for visual odometry, in *IEEE Int. Conf. Robotics and Automation (ICRA)* (Stockholm, Sweden, 2016), pp. 801–808.

[36] D. Fethi, N. Abdelkrim, L. Kahina, H. Mustapha and B. Abdelouahab, Cooperative SLAM for multiple UGVs navigation using SVSF filter, *Aautomatika J. Control Meas. Electron. Comput. Commun.* **58**(1) (2017) 119–129.

[37] S. A. Gadsden, D. Dunneb, R. Tharmarasab, S. R. Habibia and T. Kirubarajanb, Application of the smooth variable structure filter to a multi-target tracking problem, in *Proc. Signal Processing Sensor Fusion and Target Recognition* (McMaster University, Hamilton, ON, 2011) p. L8S 4L7.

[38] Y. Shenghai, W. Han and X. Lihua, Survey on localization systems and algorithms for unmanned systems, *Unmanned Syst.* **9** (2021) 129–163.

[39] A. Beauvisage, K. Ahiska and N. Aouf, Multimodal tracking framework for visual odometry in challenging illumination conditions, *2020 IEEE Int. Conf. Robot. Autom. (ICRA)* (2020), pp. 11133–11139, doi: 10.1109/ICRA40945.2020.9196891.

[40] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera and R. Toledo, Multispectral stereo odometry, *IEEE Trans. Intell. Transp. Syst.* **16**(3) (2015) 1210–1224, doi: 10.1109/TITS.2014.2354731.

[41] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys and L. V. Gool, Night-to-day image translation for retrieval-based localization, *2019 Int. Conf. Robot. Autom. (ICRA)* (New York, 2019), pp. 5958–5964, doi: 10.1109/ICRA.2019.8794387.