Appearance-based Indoor Navigation by IBVS using Mutual Information

Suman Raj Bista, Paolo Robuffo Giordano, François Chaumette

Appearance-based Indoor Navigation by IBVS using Mutual Information

Suman Raj Bista
Inria at Irisa and Inria Rennes-
Bretagne Atlantique, Rennes, France
Email: suman-raj.bista@inria.fr

Paolo Robuffo Giordano
CNRS at Irisa and Inria Rennes-
Bretagne Atlantique, Rennes, France
Email: prg@irisa.fr

François Chaumette
Inria at Irisa and Inria Rennes-
Bretagne Atlantique, Rennes, France
Email: francois.chaumette@inria.fr

Abstract—This paper presents a complete framework for image-based navigation from an image memory that exploits mutual information and does not need any feature extraction, matching or any 3D information. The navigation path is represented by a set of automatically selected key images obtained during a prior learning phase. The shared information (entropy) between the current acquired image and nearby key images is exploited to switch key images during navigation. Based on the key images and the current image, the control law proposed by [1] is used to compute the rotational velocity of a mobile robot during its qualitative visual navigation. Using our approach, real-time navigation has been performed inside a corridor and inside a room with a Pioneer 3-DX equipped with an on-board perspective camera without the need of accurate mapping and localization.

I. INTRODUCTION

Visual navigation has become mainstream for mobile robot navigation [2]. One of the common navigation approaches relies on an accurate and consistent 3D model of the environment like in simultaneous localization and mapping (SLAM) [3] based methods, where navigation can be easily performed via 3D pose-based control in a metric space [4]. In another approach, also known as appearance-based, the environment is represented by a set of key images in topological graphs [5], [6], [7], which are far less complex than metric maps. This approach works directly in the sensor space and does not need a prior 3D model of the navigation space. The navigation is based on the comparison between the current view and the images in the map that are generally based on global descriptors, like entire image [1], [8], histograms [9], [10], or image gradient [11]; or by using local descriptors, like interest points [5], [6]; line segments [7]. Using this approach, [5], [7] show that accurate mapping and localization are not mandatory for visual navigation.

Robust extraction and matching/tracking of the features over a large environment is still a bottleneck for the visual navigation schemes. However, another possibility is to directly exploit a comparison between the full key images and current image, using Fourier space [12], cross-correlation [13], [14], image intensity [15], [16], histogram [10], or mutual information [1]. In [17], Sum of Squared Difference (SSD) has been used for changing key images but the control law is based on the vanishing points of the corridor and the key images selection was not automatic. The work of [1] shows visual path following using mutual information in outdoor environment but it lacks an appropriate selection and switching

of the key images. Also, the robot should be near to the starting point in the learning. Compared to intensity-based similarity metrics, like SSD or cross-correlation, mutual information is robust to illumination variations and to large occlusions [18]. Additionally, mutual information is a classic similarity measure especially for multi-modal registration techniques in medical imaging and remote sensing [19], [20], [21]. Therefore, in this work we choose to use mutual information for navigation of a mobile robot in an indoor environment.

In our work, we propose a complete method for the indoor navigation (mapping, localization in the topological graph and motion control) based on image memory using mutual information. We have extended the work of [1] with automatic key images selection, initial localization in the map, key images switching for successive localization and use of multiple key images for control. Indeed in [1], all the images of the sequence acquired during the learning step were used, and the synchronization between the current image and the corresponding image in the sequence learned was a real challenge as adequate switching of key frames was not employed. The use of few key images, initial localization and automatic switching of key images removes the problem of synchronization and enables the robot to use different forward velocities than those used for the learning phase. The initial localization makes it possible for the robot to start from any position within the map. Without the need of any accurate mapping and localization, 3D information, and feature extraction and matching, our method is able to perform navigation in indoor environment using mutual information.

The next section describes the complete framework for mapping and navigation. Section III presents experimental results with a real robotic system. Finally, some concluding remarks are reported in Section IV.

II. OVERVIEW

A. Constraints

We consider a non-holonomic mobile robot of unicycle type equipped with a fixed and pre-calibrated perspective camera. The robot exhibits a qualitative path following behavior without any obstacle avoidance. It is therefore suitable to prefer the center of the free space during the acquisition of the learning sequence. During navigation, it is assumed that the robot

is initially inside the mapped environment. The localization outside the mapped location is out of scope of this paper.

B. Mutual Information

In our case, mutual information is the information shared by two images. For two images I and I_k, mutual information is given by the following equation [21]

$$MI(I, I_k) = H(I) + H(I_k) - H(I, I_k), \qquad (1)$$

where $H(I)$ denotes the entropy of the image I, i.e., its variability. $H(I, I_k)$ denotes the joint entropy of the images I and I_k, i.e., the joint variability of two images. By subtracting the joint variability from the variabilities, as in (1), we obtain the shared information of the two images, which is defined as mutual information. The entropy of the image can be easily obtained from its probability distribution, i.e., the normalized histogram of the image. Let n be the total number of gray levels and $p_I(i)$ be the probability of the particular gray level i. The entropy of the image I is then defined as

$$H(I) = -\sum_{i=0}^{n-1} p_I(i) \log(p_I(i)). \qquad (2)$$

Similarly, the joint entropy between I and I_k can be obtained from joint probability distribution, i.e., normalized joint histogram between two images

$$H(I, I_k) = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{II_k}(i,j) \log(p_{II_k}(i,j)), \qquad (3)$$

where $p_{II_k}(i,j)$ is the joint probability between gray level i in I and gray level j in I_k.

Equations (1-3) give the following expression for the mutual information

$$MI(I, I_k) = -\sum_{i,j} p_{II_k}(i,j) \log\left(\frac{p_{II_k}(i,j)}{p_I(j)p_{I_k}(j)}\right). \qquad (4)$$

The analytical functions for the probabilities are given by [1]

$$p_I(i) = \frac{1}{N_x}\sum_x \phi(i - I(x)) \text{ and} \qquad (5)$$

$$p_{II_k}(i,j) = \frac{1}{N_x}\sum_x \phi(i - I(x))\phi(j - I_k(x)), \qquad (6)$$

where N_x is the total number of pixels in the image and $\phi(x)$ is a function used to fill the histogram such that $\phi(\xi) = 1$ if $\xi = 0$ and 0 otherwise.

For gray images, $n = 256$. If larger number of gray level are used for the construction of the histogram, there will be more empty bins. Moreover, the histograms with larger number of bins are expensive to construct in terms of memory as well as time. So, in practice, smaller value of n (like 8, 16) are used. The other advantage of having smaller number of bins is that the cost function is smoother with a larger convergence domain [1]. For $n = N_c$ bins, the scaled image $\overline{I}(x)$ is given as

$$\overline{I}(x) = I(x)\frac{N_c - 1}{255}, \qquad (7)$$

which has no longer integer values of intensities. Thus, function ϕ in (5-6) has to be modified to use real values. The solution for this is to use a B-spline function (that corresponds to B-spline interpolation) [22].

The similarity measure given by mutual information is meaningful if the shared portions of the images are aligned while calculating the joint probability. Therefore, to measure the similarity between two images, the images are registered with each other such that mutual information between them is maximized. The maximization problem is given as

$$\hat{\rho} = \arg\max_{\rho} MI(I(\rho), I_k), \qquad (8)$$

where ρ is the transformation that maximizes (8). The mutual information in this optimal transformation $\hat{\rho}$ is given as

$$\widehat{MI}(I, I_k) = MI(I(\hat{\rho}), I_k). \qquad (9)$$

In our approach, mutual information is calculated from (8-9) using the method proposed by [18] with 8 histogram bins. Multi-scale registration [23] has been performed instead of single scale, which enables to have fast and more accurate registration. Only 2D translation has been used as the transformation function during the registration in our experiments. In other words, the output of equation (8) is a 2D translation between the images. The choice of this transformation is reasonable as we are just controlling the one degree of vertical rotational velocity that defines the heading angle of the robot.

C. Key images selection (Mapping)

In the proposed method, the environment is represented by a few number of key images that are selected automatically, unlike [1]. The key images selection procedure is sketched in Fig. 1. The first acquired image is always stored in the database as a key image (first node in the graph). Let I_c be the most recently acquired image and I_k be the most recent key image. If the mutual information metrics is smaller than a certain threshold, or the transformation is too large, I_c becomes the new I_k. Then the process continues. The second criterion is essentially required to get a sufficient number of key images while turning. The last acquired image is also stored in the database, which helps in determining when the robot has to stop at the end of the navigation.
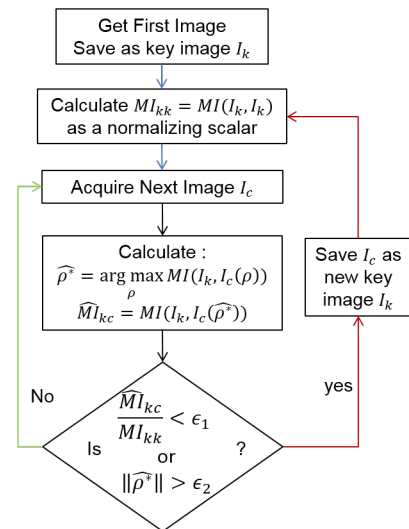

Fig. 1: Key Images selection using mutual information.

D. Initial Localization in the map

The navigation starts with the initial localization where the first acquired image (I_a) is compared with all the key images

based upon the maximization of the mutual information (8-9). The key image that has maximum mutual information with I_a is selected. Let it be I_k. Then the adjacent key image with second maximum value of mutual information is also selected. This image can be either I_{k+1} or I_{k-1}. If the robot is assumed to move along the same direction as the images arranged in the database, I_a is between I_{k-1} and I_k, or I_k and I_{k+1}. For simplicity, we denote the previous key image as I_P and the next key image as I_N. Hence, the position of I_a in the map is in between I_P and I_N as shown in Fig. 2.
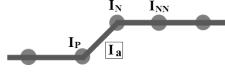

Fig. 2: Localization in the map.

E. Successive Localization (Key Images Switching)

After the initial localization in the map, further localizations can be done by just comparing with few adjacent key images. The next key image I_N and the second next key image I_{NN} are compared with the current acquired image I_a. Let $\widehat{MI}(...)$ be the mutual information between the images. Then switching of key images is done when at least one of the following criteria is fulfilled for consecutive acquired images I_a and I_{a+1}

$$
\begin{array}{ll}
\widehat{MI}(I_a, I_{NN}) & > \quad \widehat{MI}(I_a, I_N) \text{ or} \\
\dfrac{\widehat{MI}(I_a, I_{NN})}{MI(I_a, I_a)} < \epsilon & \&\& \quad \dfrac{\widehat{MI}(I_a, I_N)}{MI(I_a, I_a)} < \epsilon,
\end{array}
$$

where ϵ is a small constant (say 0.15). The denominator in the second criteria is a normalization that helps to select ϵ for a given environment. The second criterion is essentially useful when the registration error is large or the robot is lost, which may sometimes occur with regions with no significant texture. After switching the images, I_N becomes I_P, I_{NN} becomes I_N, and next key image from I_N becomes I_{NN}. Then the process repeats. When the end of the database is reached, I_{NN} will not be available and I_N will be the last image acquired during the mapping. Therefore, the navigation needs to be stopped. Otherwise, the robot will be moving out of the mapped environment. The coarse registration (with few iterations say 3-4) using (8-9) is found to be sufficient for key images switching from our experiments.

F. Control

For navigation, the robot is not required to accurately reach each key image of the path, or to accurately follow the learned path. In practice, the exact motion of the robot should be controlled by an obstacle avoidance module [24]. Therefore, translational velocity is kept constant and reduced to a smaller value when turning. Such turnings are automatically detected by looking at the commanded rotational velocity. The rotational velocity is derived using key images and current image within an image-based visual servoing (IBVS) control law [25].

Let us define a vector of visual features as $\mathbf{s}$, the camera velocity expressed in camera frame as $\mathbf{u_c} =$ $(v_{cx}, v_{cy}, v_{cz}, \omega_{cx}, \omega_{cy}, \omega_{cz})$, where v is the linear velocity and ω is the rotational velocity around the given axes. The velocity of $\mathbf{s}$ can be related via an interaction matrix $\mathbf{J_s}$ [25] to $\mathbf{u_c}$ as

$$\dot{\mathbf{s}} = \mathbf{J_s u_c}. \tag{10}$$

During visual navigation, the robot moves so as to maximize mutual information with the current reference image. Comparing with classical visual servoing, the gradient of mutual information ($\mathbf{L_{MI}}$) gives the error term whereas the Hessian of mutual information ($\mathbf{H_{MI}}$) gives the interaction matrix $\mathbf{J_s}$. The derivatives of mutual information can be calculated by following [1]

$$\mathbf{L_{MI}} = \sum_{i,j} \mathbf{L_{p_{II_k}}} \left(1 + \log \left(\frac{p_{II_k}}{p_{I_k}} \right) \right), \tag{11}$$

$$
\begin{aligned}
\mathbf{H_{MI}} = & \sum_{i,j} \mathbf{L}^T_{p_{II_k}} \mathbf{L_{p_{II_k}}} \left(\frac{1}{p_{II_k}} - \frac{1}{p_{I_k}} \right) \\
& + \mathbf{H_{p_{II_k}}} \left(1 + \log \left(\frac{p_{II_k}}{p_{I_k}} \right) \right).
\end{aligned} \tag{12}
$$

If N_x is the number of pixels considered in the images and ϕ is a twice differentiable B-spline function, the derivatives of the joint probability using $Nc\,(< 255)$ grey levels are as follows

$$p_{II_k}(i,j) = \frac{1}{N_x} \sum_x \phi(i - \bar{\mathbf{I}}(\mathbf{x})) \phi(j - \overline{\mathbf{I_k}}(\mathbf{x})), \tag{13}$$

$$\mathbf{L_{p_{II_k}}}(i,j) = \frac{1}{N_x} \sum_x \mathbf{L}_{\phi(i - \bar{\mathbf{I}}(\mathbf{x}))} \phi(j - \overline{\mathbf{I_k}}(\mathbf{x})), \tag{14}$$

$$\mathbf{H_{p_{II_k}}}(i,j) = \frac{1}{N_x} \sum_x \mathbf{H}_{\phi(i - \bar{\mathbf{I}}(\mathbf{x}))} \phi(j - \overline{\mathbf{I_k}}(\mathbf{x})), \tag{15}$$

where,

$$\mathbf{L}_{\phi(i - \bar{\mathbf{I}}(\mathbf{x}))} = -\frac{\partial \phi}{\partial i} (\nabla \bar{\mathbf{I}} \mathbf{L_x}), \tag{16}$$

$$\mathbf{H}_{\phi(i - \bar{\mathbf{I}}(\mathbf{x}))} = \frac{\partial^2 \phi}{\partial i^2} (\nabla \bar{\mathbf{I}} \mathbf{L_x})^T (\nabla \bar{\mathbf{I}} \mathbf{L_x}) - \frac{\partial \phi}{\partial i} (\mathbf{L_x^T} \nabla^2 \bar{\mathbf{I}} \mathbf{L_x} + \nabla \bar{\mathbf{I}} \mathbf{H_x}), \tag{17}$$

where $\nabla \bar{\mathbf{I}} = \begin{bmatrix} \nabla_x \bar{\mathbf{I}} & \nabla_y \bar{\mathbf{I}} \end{bmatrix}$ are the image gradients, $\nabla^2 \bar{\mathbf{I}} = \begin{bmatrix} \nabla_{xx} \bar{\mathbf{I}} & \nabla_{xy} \bar{\mathbf{I}} \\ \nabla_{yx} \bar{\mathbf{I}} & \nabla_{yy} \bar{\mathbf{I}} \end{bmatrix}$ are the gradients of image gradient, $\mathbf{L_x}$ is the interaction matrix of a point that links its displacement in the image plane to the camera velocity, and $\mathbf{H_x}$ is the Hessian of the point w.r.t. the camera velocity. The complete expression for $\mathbf{L_x}$ and $\mathbf{H_x}$ for six degrees of freedom is given in [26].

For the considered unicycle-like robot (Fig. 3), $\mathbf{u_c}$ can be
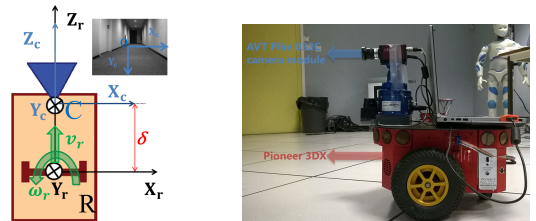

Fig. 3: Top view of robot (orange) equipped with a perceptive camera (blue) with its optical axis perpendicular to axis of robot rotation (left) and experimental platform (right).

expressed in terms of the robot velocity $\mathbf{u} = (v_r, \omega_r)$ as

$$\mathbf{u_c} = (-\delta \omega_r, \quad 0, \quad v_r, \quad 0, \quad -\omega_r, \quad 0), \tag{18}$$

where δ is the distance between the camera center and the robot center of rotation, v_r is the forward velocity and ω_r is the rotational velocity of the robot. Now, from (10) and (18), we obtain

$$\dot{\mathbf{s}} = \mathbf{J}_v v_r + \mathbf{J}_\omega \omega_r, \qquad (19)$$

where $\mathbf{J}_v$ and $\mathbf{J}_\omega$ are the Jacobian associated with v_r and ω_r respectively. In order to maximize mutual information, we set v_r as constant and control ω_r as [25]

$$\omega_r = -\mathbf{J}_\omega^+(\lambda(\mathbf{s} - \mathbf{s}^*) + \mathbf{J}_v v_r), \qquad (20)$$

where λ is a positive gain, $\mathbf{J}_\omega^+$ is the pseudo-inverse of $\mathbf{J}_\omega$ and $(\mathbf{s} - \mathbf{s}^*)$ is the error term. $\mathbf{J}_v$ and $\mathbf{J}_\omega$ are the Hessian of mutual information w.r.t v_r and ω_r respectively whereas $(\mathbf{s} - \mathbf{s}^*)$ is the gradient of the mutual information. Since we are controlling only ω_r, in (20),

$$(s - s^*) = L_{MI\omega}, \quad J_v = \frac{\partial L_{MI\omega}}{\partial v_r} = H_{MIv},$$
$$\text{and} \qquad J_\omega = \frac{\partial L_{MI\omega}}{\partial \omega_r} = H_{MI\omega}, \qquad (21)$$

where $L_{MI\omega}$ is the Jacobian of mutual information w.r.t. ω_r, and H_{MIv} and $H_{MI\omega}$ are the Hessian of the mutual information w.r.t. v_r and ω_r respectively. They can be derived as follows:

Using (18), and the interaction matrix of point that links its displacement w.r.t. to camera velocity (from [26]), we obtain

$$\mathbf{L}_{\mathbf{xu}\omega} = \begin{bmatrix} -\left(\frac{\delta}{Z} + 1 + x^2\right) \\ -xy \end{bmatrix} \simeq \begin{bmatrix} -(1 + x^2) \\ -xy \end{bmatrix} \text{ for } \delta \ll Z, \qquad (22)$$

where $\mathbf{L}_{\mathbf{xu}\omega}$ is the interaction matrix of the point that links its displacement in the image plane to ω_r, and δ can be safely neglected with respect to Z (depth of point from image plane). Hence, from (11-16) using (22) for $\mathbf{L_x}$, we can obtain $L_{MI\omega}$, which is a scalar. Differentiating (22) w.r.t. v_r and ω_r, we obtain $\mathbf{H_{xuv}}$ and $\mathbf{H_{xu\omega}}$, which are the Hessian matrix of the point that links its displacement in the image plane to v_r and ω_r respectively. Therefore, we obtain

$$\mathbf{H_{xu\omega}} = \begin{bmatrix} 2x(1 + x^2) \\ y(1 + 2x^2) \end{bmatrix} \text{ and } \mathbf{H_{xuv}} = \begin{bmatrix} -\frac{2x^2}{Z} \\ -\frac{2xy}{Z} \end{bmatrix} \simeq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (23)$$

since x^2 and xy can be safely neglected w.r.t. Z, where (x, y) is the normalized point in the image plane. From (11-17) using (22-23) for $\mathbf{L_x}$ and $\mathbf{H_x}$ respectively in (16-17), we obtain H_{MIv} and $H_{MI\omega}$, which are scalar. As $H_{MIv} = 0$, (21) becomes

$$(s - s^*) = L_{MI\omega}, \ J_v \simeq 0, \text{ and } J_\omega = H_{MI\omega}. \qquad (24)$$

Since visual servoing is known to be robust against modeling errors [25], such approximations in (22-24) are reasonable. Thus, from (20) and (24), the final expression for ω_r is obtained as

$$\omega_r = -\frac{\lambda L_{MI\omega}}{H_{MI\omega}}, \qquad (25)$$

which is similar to the control scheme proposed in [1]. In order to smooth the rapid steering actions when switching between frames, a feed-forward command is also added to ω_r. The final expression for the rotational velocity is calculated by using I_N and I_{NN} as follows:

$$\omega_r = -\lambda\left(h_1 \frac{L_{MI\omega(I_a,I_N)}}{H_{MI\omega(I_a,I_N)}} + h_2 \frac{L_{MI\omega(I_a,I_{NN})}}{H_{MI\omega(I_a,I_{NN})}}\right), \qquad (26)$$

where h_1 and h_2 are positive weights such that $h_1 + h_2 = 1$, $L_{MI\omega(I_a,I_N)}$ and $H_{MI\omega(I_a,I_N)}$ are derivatives of mutual information between I_a and I_N, and $L_{MI\omega(I_a,I_{NN})}$ and $H_{MI\omega(I_a,I_{NN})}$ are derivatives of mutual information between I_a and I_{NN}.

Thus, our complete framework uses only the information directly obtained from image without any feature extraction or 3D information. From this information, we derive the required rotational velocity using *IBVS*, which makes the robot to follow the learned path successfully without any need for accurate mapping or localization.

III. EXPERIMENTAL RESULTS

The experiments were performed with a Pioneer 3-DX equipped with AVT Pike 032C camera module. The image resolution in the experiments was 640×480. All computations, except for the low-level control, were performed on a laptop with 3-GHz Intel Core i7-3540M CPU. The mapping was done off-line, whereas the navigation experiment was performed on-line at 5 Hz. The image coordinates have been normalized by the camera intrinsic parameters before deriving the rotational velocity. Eight histogram bins and fourth order B-spline functions have been used for experiments. The experiments have been performed in an indoor environment, i.e., inside a room and a corridor with $\lambda = 1.0$, $h_1 = 0.7$, and $h_2 = 0.3$ in (26). Even though the proposed method has been validated via a simple navigation path with linear and curved trajectories, the method can be easily extended for graphs with intersections and multiple paths. The trajectories presented below have been obtained from the odometry of the Pioneer, which clearly shows that the robot approximately follows the learned path. Since our navigation is qualitative rather than quantitative, the odometry obtained from Pioneer 3-DX is highly accurate enough to serve our purpose.

A. Mapping

1278 images have been acquired as learning sequence inside the robotics room. 10 images shown in Fig. 4 have been selected automatically from the mapping procedure described in Sect. II-C as key frames. Similarly, 36 images as shown in Fig. 7 have been selected as key images from 3881 images acquired for the 20m path consisting of a robotics room and a corridor. The trajectories obtained from the odometry are shown by a red curve in Figs. 5-9, where the red symbol $*$ represents the location of the key images. The obtained key images are able to represent the learned path. There are more key images over a small distance in case of turnings and of locations where the surface does not have good textures.
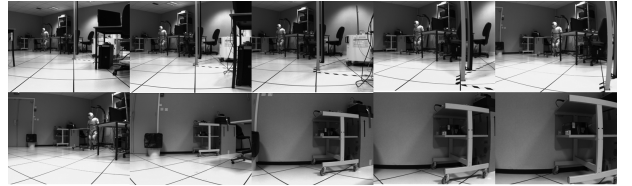

Fig. 4: Key images for navigation inside the robotics room.

B. Navigation

The robot was placed inside the mapped environment with the camera facing towards the mapped direction (Initial position shown by green dot). The forward velocity was set to 0.08m/s and reduced to 0.05m/s when turning, whereas the rotational velocity was controlled by the navigation algorithm. During navigation, the robot has been able to follow the learned trajectory as shown by the blue curve in Figs. 5-9, with automatic switching of the key images.

1) Navigation inside a robotics room: Here we performed two experiments: the first one to validate our approach and the second one to show the robustness of our method in case of changes in illumination. In both cases, navigation has been performed with some changes in environment, like the table that is shown in the last key images in Fig. 4, was moved during the navigation. Fig. 5 shows the navigation of the Pioneer in the map without any change in illumination. The navigation in presence of change of illumination is depicted in Fig. 6, where all lights have been made dimmer. In both cases, our framework has been able to select the initial key frames and follow the learned path.
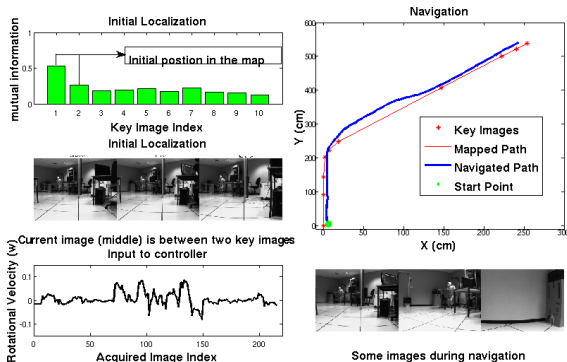

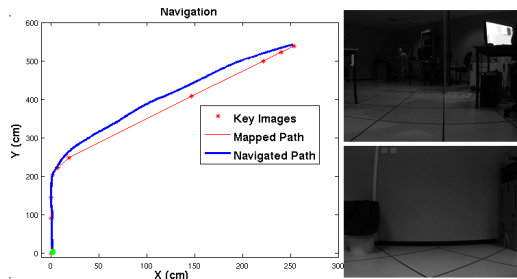Fig. 5: Initial localization and navigation inside the robotics room.

Fig. 6: Navigation inside the robotics room with the change in illumination.

2) Navigation in a room and a corridor: Fig. 8 shows the navigation in a room and a corridor where the path consists of multiple turns. The robot moved from inside the room to the corridor. The robot followed the learned path with turning whenever it was required despite presence of moving people during navigation. Right angle turning is a challenging task especially in the corridors with similar/low texture. However, the key images obtained from the mapping part were still
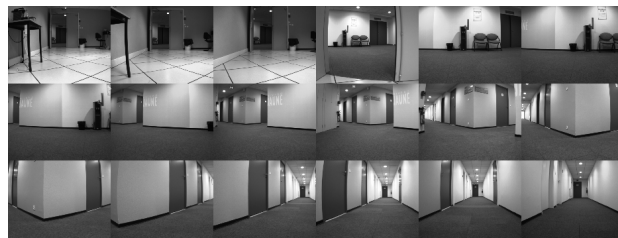

Fig. 7: Even key images ($2^{nd}, 4^{th}, ...$) of the room and corridor.

able to handle such situations. The lateral drift was within 7 cm from the mapped position. Indeed, our objective was just to perform a successful navigation without any accurate pose correction with the learned path.
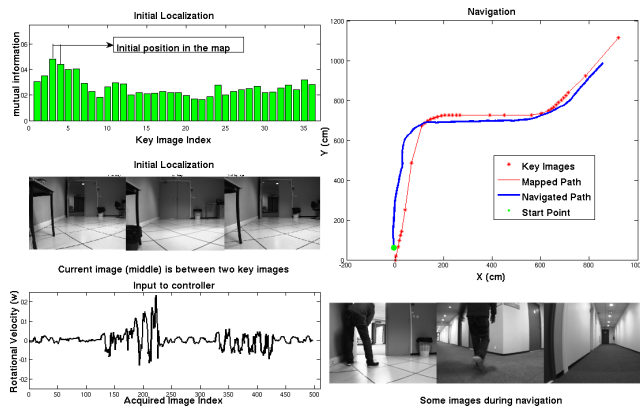

Fig. 8: Navigation between the room and the corridor.

3) Navigation in a corridor with peoples passing by: In this case, we performed the navigation in two corridors of length 30m and 31m. They are represented by 136 and 95 key images respectively, which were selected from the learning sequence of 6046 and 7507 images respectively. The second corridor consists of large windows that allow seeing outdoor so that robot undergoes change in illumination as it enters inside. Even in presence of moving people and change in illumination (Fig. 9), the robot was able to follow the learned trajectory.
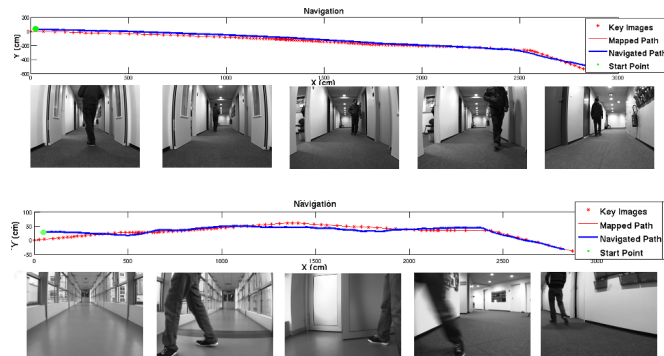

Fig. 9: Navigation in the corridors.

C. Discussion

The presented results show the viability of our approach in different scenarios and constraints. The robot has been able to

follow autonomously the learned path from the start position. Our navigation framework is based upon the maximization of the mutual information, where the robot performs the navigation task to maximize the mutual information with the nearest key images with one degree of freedom. Based on this, the key images are switched automatically and the appropriate rotational velocity is set for allowing the robot to follow the learned path. IBVS has been able to keep the error within small bounds. The robot did not exactly follow the learned path as we are controlling just one degree of freedom without any 3D pose correction, and also partly due to the approximation of the path by straight lines. Still, successful navigation has been performed. Our initial localization is based upon the fact that the nearby key images share more entropy than the farther ones. The bar graphs in the Figs. 5 and 8 confirm this idea. Initial localization is key for successful navigation. Therefore, the images are registered at finer level before calculating mutual information. This process is time consuming. However, use of multiple pyramids (3 have been used in our experiments) and the use of translation transformation allow speeding up the process. After initial localization, the successive navigation has been done in real time at 5Hz.

Comparing with feature based methods [5], [7], our method performs better especially where reliable point/lines based features cannot be detected, resulting in failure in tracking/matching because the mutual information uses the entire image information and does not require feature detection. Besides that, mutual information is also able to handle changes in lightning conditions. However, our framework also has some limitations that are mainly due to poor conditioning of the mutual information especially in cases with few or no texture at all. Nevertheless, most of these problems can be greatly avoided by selecting a proper trajectory during the mapping.

IV. CONCLUSIONS

We have presented a framework for indoor qualitative mapping and navigation based on image memory using mutual information. Our navigation is exclusively based on image information (entropy) without relying on any 3D reconstruction, feature extraction, matching or tracking process. We also showed that image-based navigation could be performed without accurately tracking the trajectory used in the learning phase, which is possible because of the topological representation of the environment and the robustness of IBVS. Using the information from the entire image makes navigation possible despite some level of occlusions (like people moving), blurs in the image, and changes in lighting conditions. Difficult situations include featureless areas like smooth/texture-less walls and rapid turnings. The rotational velocity was fed to the robot without any post processing and filtering, which sometimes resulted in non-smooth motion. More precise image processing and/or robot control strategies, like using vehicle kinematics and filtering the velocity, can address these issues. Besides this, incorporating obstacle avoidance similar to [24] will constitute a future work.

REFERENCES

[1] A. Dame and E. Marchand, "Using mutual information for appearance-based visual path following," *RAS*, vol. 61, pp. 259–270, 2013.

[2] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: A survey," *J. of IRS*, vol. 53, no. 3, pp. 263–296, 2008.

[3] A. Davison, I. Reid, N. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE T-PAMI*, vol. 29, pp. 1052–1067, June 2007.

[4] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *IJCV*, vol. 74, no. 3, pp. 237–260, 2007.

[5] S. Segvic, A. Remazeilles, A. Diosi, and F. Chaumette, "A mapping and localization framework for scalable appearance-based navigation," *CVIU*, vol. 113, pp. 172 – 187, 2009.

[6] O. Booij, B. Terwijn, Z. Zivkovic, and B. Krose, "Navigation using an appearance based topological map," in *IEEE ICRA, 2007*, pp. 3927–3932, Apr. 2007.

[7] S. Bista, P. Giordano, and F. Chaumette, "Appearance-based indoor navigation by IBVS using line segments," *IEEE Robotics and Automation Letters*, vol. 1, pp. 423–430, Jan. 2016.

[8] F. Labrosse, "Short and long-range visual navigation using warped panoramic images," *RAS*, vol. 55, pp. 675 – 684, 2007.

[9] F. Werner, J. Sitte, and F. Maire, "Visual topological mapping and localisation using colour histograms," in *ICARV, 2008*, pp. 341–346, Dec. 2008.

[10] Q. Bateux and E. Marchand, "Histograms-based visual servoing," *IEEE Robotics and Automation Letters*, vol. 2, pp. 80–87, Jan. 2017.

[11] J. Kosecka, L. Zhou, P. Barber, and Z. Duric, "Qualitative image based localization in indoors environments," in *IEEE Conf. on CVPR, 2003*, vol. 2, pp. II–3–II–8 vol.2, June 2003.

[12] A. M. Zhang and L. Kleeman, "Robust appearance based visual route following for navigation in large-scale outdoor environments," *IJRR*, vol. 28, no. 3, pp. 331–356, 2009.

[13] J. Courbon, Y. Mezouar, and P. Martinet, "Autonomous navigation of vehicles from a visual memory using a generic camera model," *IEEE Trans. Intell. Transport. Sys.*, vol. 10, pp. 392–402, Sept. 2009.

[14] Y. Matsumoto, M. Inaba, and H. Inoue, "Visual navigation using view-sequenced route representation," in *IEEE ICRA, 1996*, vol. 1, pp. 83–88 vol.1, Apr. 1996.

[15] C. Collewet, E. Marchand, and F. Chaumette, "Visual servoing set free from image processing," in *IEEE ICRA, 2008*, pp. 81–86, May 2008.

[16] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE T-RO*, vol. 27, pp. 828–834, Aug. 2011.

[17] R. F. Vassallo, H. J. Schneebeli, and J. Santos-Victor, "Visual servoing and appearance for navigation," *RAS*, vol. 31, no. 1, pp. 87–97, 2000.

[18] A. Dame and E. Marchand, "Second-order optimization of mutual information for real-time image registration," *IEEE T-IP*, vol. 21, pp. 4190–4203, Sept. 2012.

[19] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE T-MI*, vol. 22, pp. 986–1004, Aug. 2003.

[20] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Proc. SPIE: Medical Imaging 2001*, pp. 1609–1620, 2001.

[21] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *IJCV*, vol. 24, pp. 137–154, 1997.

[22] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE T-MI*, vol. 16, pp. 187–198, Apr. 1997.

[23] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "A multiscale approach to mutual information matching," in *Proc. SPIE Medical Imaging: Image Proces. Volume 3338*, pp. 1334–1344, 1998.

[24] A. Cherubini and F. Chaumette, "Visual navigation of a mobile robot with laser-based collision avoidance," *IJRR*, vol. 32, no. 2, pp. 189–205, 2013.

[25] F. Chaumette and S. Hutchinson, "Visual servo control, part i: Basic approaches," *IEEE RAM*, vol. 13, pp. 82–90, Dec. 2006.

[26] J. Lapreste and Y. Mezouar, "A hessian approach to visual servoing," in *Proc. IEEE Conf. on IROS, 2004*, vol. 1, pp. 998–1003 vol.1, Sept. 2004.