

# A new information theoretic approach for appearance-based navigation of non-holonomic vehicle

A. Dame, E. Marchand

► **To cite this version:**

A. Dame, E. Marchand. A new information theoretic approach for appearance-based navigation of non-holonomic vehicle. IEEE Int. Conf. on Robotics and Automation, ICRA'11, 2011, Shanghai, China, China. pp.2459-2464, 2011. <hal-00639679>

**HAL Id: hal-00639679**

**<https://hal.inria.fr/hal-00639679>**

Submitted on 9 Nov 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new information theoretic approach for appearance-based navigation of non-holonomic vehicle

Amaury Dame, Eric Marchand

**Abstract**—In this paper we propose a new way to achieve a navigation task for a non-holonomic vehicle. We consider an image-based navigation process. We show that it is possible to navigate along a visual path without relying on the extraction, matching and tracking of geometric visual features such as keypoint. The new proposed approach relies directly on the information (entropy) contained in the image signal. We show that it is possible to build a control law directly from the maximisation of the shared information between the current image and the next key image in the visual path. The shared information between those two images are obtained using mutual information that is known to be robust to illumination variations and occlusions. Moreover the generally complex task of features extraction and matching is avoided. Both simulations and experiments on a real vehicle are presented and show the possibilities and advantages offered by the proposed method.

## I. INTRODUCTION

In recent years, robot localization and navigation have made considerable progress. Navigation can be seen as the ability for a robot to move autonomously from an initial position to a desired one (which may be far away from the initial one). Thanks to sensor based navigation, we have seen autonomous robots in various challenging areas (from highways to deserts and even on Mars). Nevertheless the design of these autonomous robots usually relies on more than one sensor (camera, stereo sensors, lidar, GPS,...). In this paper, we propose a new method that demonstrates the capability of a mobile robot to navigate autonomously using the information provided by a monocular camera. Furthermore we will show that the proposed approach does not require any tracking nor matching process which is usually a bottleneck for the development of such approach.

Most navigation approaches consider a (partial) 3D reconstruction of the environment, leading to SLAM-like techniques. Such solutions are attractive, since the navigation task will be achieved using a classical pose-based control of the robot in the metric space. Within this context, during a learning step the environment is reconstructed using bundle adjustment approaches [16] or Kalman/particle filters based approaches [4]. Despite the complexity of the underlying problem, SLAM has proved to be a viable solution to create accurate maps of the environment [4], [17] even in large ones [10]. In this context, the control of the robot during the navigation task is a well known problem and

the main difficulties here are i) the complexity of the initial reconstruction step and ii) the matching of visual features observed during the learning step with current observations. With a monocular camera as unique sensor, these are mainly computer vision issues.

Another class of techniques relies on the definition of a visual path: the appearance-based approaches [3], [6], [15], [18]. The trajectory is no longer described in the metric space but as a set of reference images. A 2D visual servoing step allows the robot to navigate from its current position to the next key images. When the robot gets close to this image, a new key image is selected. In this context, the environment can be modeled by a graph whose nodes are the key images. A visual path in the environment is nothing but a path in the graph [15]. Working directly in the sensor space, such approaches do not require prior 3D reconstruction step. In some cases, partial reconstruction has to be considered. In [18] [8] a part of the epipolar geometry that links the current and key images is considered in order to predict the location of currently not visible features and ensure a robust tracking. In [6] homography computation wrt. the reference images allows to precisely localize the robot. In any case, the learning step of these appearance-based approaches is far less complex since it does not require any prior 3D reconstruction.

Nevertheless at navigation level, for both pose-based or image-based visual navigation, features have to be extracted or tracked in the image stream and matched with either the 3D database or key images to design the control law. Robust extraction and real-time spatio-temporal tracking or matching of these visual cues are non trivial tasks and also one of the bottlenecks of the expansion of visual navigation. In [5], it has been shown that no other information than the image intensity (the pure image signal) need to be considered to control the robot motion and that these difficult tracking and matching processes can be totally removed. Although very efficient, this approach is sensitive to light variations and, thus, can hardly be considered in outdoor environment. In this paper, we propose a new approach that no longer relies on geometrical features nor on pixels intensity [5] but uses directly the information (entropy) contained in the image signal as proposed in [7]. More precisely we will consider mutual information [19], [21]. Being closer to the signal, we will show that this approach is robust to very important illumination variations and robust to large occlusions. We show that it is possible to compute the interaction matrix that relates the variation of the mutual information to the vehicle velocity leading to the definition of the control law.

Amaury Dame is with CNRS, IRISA, INRIA Rennes Bretagne-Atlantique, Lagadic research group, France, email: Amaury.Dame@irisa.fr. Eric Marchand is with Université de Rennes 1, IRISA UMR 6074, INRIA Rennes-Bretagne Atlantique, Lagadic research group, France, email: Eric.Marchand@irisa.fr. This work is supported by DGA under contribution to student grant.

Let us emphasize that since mutual information is computed from the whole images (current and key image) it is possible to directly control the motion of the vehicle along a given path without any feature extraction or matching. Furthermore no 3D reconstruction of the environment is necessary.

We will demonstrate the efficiency of this new approach on a navigation task carried out at 0.5 m/s over 400 meters. Images are acquired at 30Hz (nearly 25.000 images where acquired and processed in real-time during this navigation task).

## II. NAVIGATION PROCESS OVERVIEW

In this work, we consider a non-holonomic robot with a camera mounted on the front. Our goal is not to localize the robot within its environment (visual odometry) but only to ensure that it is able to reproduce a visual path defined as a set of images previously acquired by the camera.

### A. Learning step: definition of the visual path

With respect to previous approaches that rely on 3D reconstruction (eg, [16]) or even on appearance-based approaches [18], the learning step of the approach is simple. It does not require any feature extraction nor scene reconstruction: no image processing is done, only raw images are stored. The vehicle is driven manually along a desired path. While the vehicle is moving, the images acquired by the camera are stored chronologically thus defining a trajectory in the image space. Let us call  $I_0^*, \dots, I_N^*$  the key images that define this visual path.

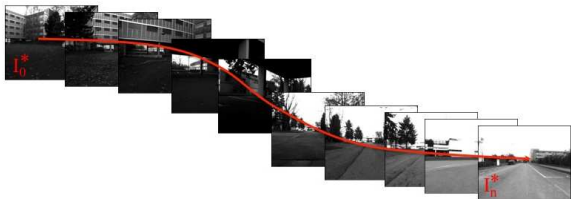


Fig. 1. Key images that define the visual path. This visual path is learned prior to the navigation step.

### B. Navigation step: following the visual path

The vehicle is initially positioned close to the initial position of the learned visual path (defined by the image  $I_0^*$ ). The navigation is performed using a visual servoing task. Figure 2 shows the general control scheme used for the navigation. In [3], [16], [18] the considered control scheme are either pose-based control law or consider classical visual servoing process based on the use of visual features extracted from the current and key images ( $I$  and  $I_k^*$ ).

In this work the definition of a new control law is proposed. One of the originality of this work is that, rather than relying on features extraction and tracking, we build the control law directly from the information shared by  $I$  and  $I_k^*$  the mutual information [19]. When the mutual information between two images is maximized, the two images are similar. We then control the robot in order to

maximize the mutual information between  $I$  and  $I_k^*$ . As for any visual servoing scheme it is then necessary to exhibit the Jacobian that links the variation of the mutual information to the control input of the robot (that is the steering angle  $\psi$  or the camera rotational velocity  $\dot{\rho}$ ) needed to follow the path with a constant translational velocity  $v$ . This process is presented in the next section. In the same time, when the vehicle reaches the neighboring key image  $I_k^*$ , a new one  $I_{k+1}^*$  is selected in the visual path. To achieve a seamless switching between key images, a specific process described in section IV is proposed.

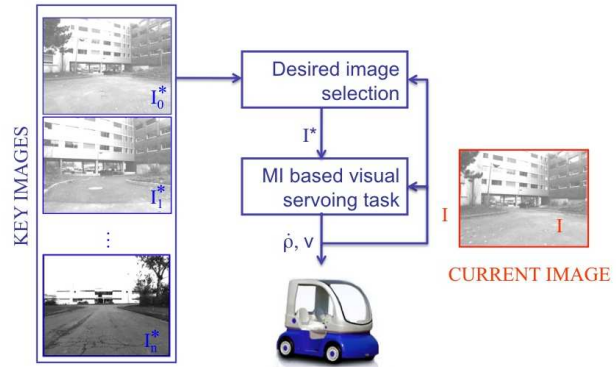


Fig. 2. Navigation based on multiple visual servoing tasks.

## III. MUTUAL INFORMATION BASED NAVIGATION

In [7], it has been shown that it is possible to achieve 6 DOF visual servoing task using only information contained in the images acquired from the camera mounted on a robot and one reference image. The desired position of the robot is reached by maximizing the mutual information between the two images. Since mutual information is robust to illumination variations and occlusions, the use of mutual information-based visual servoing is well suited for outdoor navigation tasks.

### A. Mutual information

In this section, a brief reminder of the definition of mutual information is given. Mutual information is the information shared by two signals (here, images). For the two signals  $X$  and  $Y$ , mutual information is given by the following equation [19]:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \quad (1)$$

where  $H(X)$  denotes the entropy of the signal  $X$ , that means its variability.  $H(X, Y)$  denotes the joint entropy of the signals  $X$  and  $Y$ , that is the joint variability of the system defined by the two signals. By subtracting the joint variability from the variabilities, as in equation 1, we obtain the shared information of the two signals that is mutual information.

In the present work we focus on mutual information between two images. The desired image (in the navigation context, desired image is the current key image of the visual

path) is noted  $I^*$  and the current image acquired by the vehicle camera is noted  $I$ . The original definition of mutual information given in [19], [21] can thus be used as:

$$MI(I, I^*) = \sum_{r,t} p_{rt}(r, t) \log \left( \frac{p_{rt}(r, t)}{p_r(r)p_t(t)} \right). \quad (2)$$

$p_{rt}$ ,  $p_r$  and  $p_t$  are the probabilities involved in the computation of the entropy.  $p_r(r)$  is the probability that a pixel of the desired image  $I^*$  has a value  $r$ ,  $p_t(t)$  is the probability that a pixel of the current image  $I$  has a value  $t$ , and the joint probability  $p_{r,t}(r, t)$  is the probability that the couple of value  $(I^*, I)$  at one pixel position has a value  $(r, t)$ . These probabilities are obtained by normalizing the histograms and joint histogram of the images. Their analytical formulations are given by:

$$p_{rt}(r, t) = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi(r - I^*(\mathbf{x})) \phi(t - I(\mathbf{x})) \quad (3)$$

$$p_r(r) = \sum_t p_{rt}(r, t) = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi(r - I^*(\mathbf{x})) \quad (4)$$

$$p_t(t) = \sum_r p_{rt}(r, t) = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi(t - I(\mathbf{x})) \quad (5)$$

where  $N_{\mathbf{x}}$  is the number of points  $\mathbf{x}$  in the region of interest (the complete image in our case).  $\phi(\xi)$  is the function used to fill the histogram. Typically  $p_r(r)$  is incremented each time  $I^*(\mathbf{x}) = r$ . Then  $\phi(\xi) = 1$  if  $\xi = 0$  and null otherwise. The equation of  $p_t(t)$  is similar to the one of  $p_r(r)$ .

Since the two images are typical 8 bits images with 256 gray level values, the initial definition of mutual information is given for 256 entries for  $r$  and  $t$ . This definition gives a cost function that is subject to noise, artifacts and then local extrema that may induce issues in a non-linear optimization process [13].

Several modifications on the computation allow to have a smooth cost function with a large convergence domain. The first is to consider a smaller number of entries for each histograms [22]. The effect is to smooth the extremum and enlarge the convergence domain. To do so, image intensities used in equation (3), (4) and (5) are scaled to fit in the new number of bin  $N_c$ . Let us note  $\bar{I}$  and  $\bar{I}^*$  respectively the scaled images  $I$  and  $I^*$ :

$$\bar{I}(\mathbf{x}) = I(\mathbf{x}) \frac{N_c - 1}{255} \quad \bar{I}^*(\mathbf{x}) = I^*(\mathbf{x}) \frac{N_c - 1}{255}. \quad (6)$$

The intensities of this images are no more integer values. Thus, the original  $\phi$  function has to be modified to update the histograms entries using real values (e.g. to compute  $\phi(r - \bar{I}^*(\mathbf{x}))$ ). A solution of this problem is given by the Partial Volume Interpolation [12] that defines  $\phi$  as a first order B-spline (corresponding to a bilinear interpolation).

An other operation is obtained by improving this interpolation. Instead of using a simple bilinear interpolation to compute the histograms ( $\phi = B_1$ ), B-splines of higher order are used [7]. In the present work, we consider  $\phi = B_3$  (this will also be necessary to compute the second order derivatives of mutual information).

Finally, the images  $I$  and  $I^*$  can be smoothed (using for example a gaussian filter) to increase the domain of convergence by smoothing mutual information [14]. Moreover an interesting effect of the image filtering is a convexification of the cost function. The filter variance has however to be controlled to keep the accuracy of the maximum.

Combining the three operations, the obtained MI is smooth with a wide and accurate maximum and thus adapted for the optimization problem that will be used in the next section.

## B. Navigation using Mutual information

1) *Visual Servoing using Mutual information*: In the general visual servoing formulation the goal is to minimize a dissimilarity measure (generally the difference) between some desired and current feature using a non linear minimization [2]. Such approaches have been already used in navigation [3], [18]. Our goal is to propose a more direct approach that uses the image as a whole and that does not rely on geometric features and, hence, avoid the tracking and matching steps. Rather than minimizing, as usual, the error between current and desired features, the goal is to maximize the amount of information shared by the two images. Considering that our vehicle has a constant translation velocity  $\mathbf{v}$ , we will control only the vehicle steering angle  $\rho$ . The navigation task toward the next key image  $I^*$  can be seen as an optimization process where the cost function is defined as the mutual information between  $I$  and  $I^*$  wrt. the angle  $\rho$ :

$$\hat{\rho}^* = \arg \max_{\rho} f(\rho) \quad \text{with} \quad f(\rho) = MI(I^*, I(\rho)). \quad (7)$$

This maximization is performed by updating the parameter  $\rho$  to find a null derivative of mutual information using a non linear optimization. Using the parameter update  $\delta\rho$ , the expression to maximize can typically be rewritten as :

$$MI(I^*, I(\rho_{t+1})) = MI(I^*, I(\rho_t + \delta\rho)). \quad (8)$$

where  $\rho_t$  is the current steering angle and  $\rho_{t+1}$  is the steering angle at the next iteration. As it is demonstrated in [1], this formulation is equivalent to the inverse compositional formulation, where the expression to maximize is the following one:

$$f = MI(I^*(-\delta\rho), I(\rho_t)). \quad (9)$$

As it will be explained later, this formulation allows to precompute some terms and then have a faster computation. Using a classical Newton's method, the parameter update can be computed using:

$$\delta\rho = -\alpha \ddot{f}^{-1} \dot{f} \quad (10)$$

where  $\dot{f}$  and  $\ddot{f}$  are respectively the derivative and second derivative of the  $f$  function wrt.  $\rho$ .  $\alpha \in [0, 1]$  is a scalar factor that allows to set the speed of the convergence (in our navigation task the goal is not to get to the maximum in one iteration).

Since  $I^*$  is now depending on  $\rho$ , the expressions defined in equations (4) and (3) depends on  $\rho$  too. The derivatives

of mutual information are then given by the following equations:

$$\dot{f} = \frac{\partial MI(I, I^*(\rho))}{\partial \rho} = \sum_{r,t} \frac{\partial p_{rt}}{\partial \rho} \left( 1 + \log \left( \frac{p_{rt}}{p_r} \right) \right) \quad (11)$$

$$\begin{aligned} \ddot{f} &= \frac{\partial^2 MI(I, I^*(\rho))}{\partial \rho^2} \\ &= \sum_{r,t} \frac{\partial p_{rt}^\top}{\partial \rho} \frac{\partial p_{rt}}{\partial \rho} \left( \frac{1}{p_{rt}} - \frac{1}{p_r} \right) + \frac{\partial^2 p_{rt}}{\partial \rho^2} \left( 1 + \log \frac{p_{rt}}{p_r} \right) \end{aligned} \quad (12)$$

Although it is classical to consider the second term of (12) as null [9], [20], in this work the computation of the exact second derivative of mutual information is used. The two previous expressions depend on the joint probability derivatives. Using equation (3), the formulation of the derivatives gives:

$$\frac{\partial p_{rt}}{\partial \rho} = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \frac{\partial \phi}{\partial \rho} (r - \bar{I}^*(\mathbf{x}, \rho)) \phi(t - \bar{I}(\mathbf{x})) \quad (13)$$

$$\frac{\partial^2 p_{rt}}{\partial \rho^2} = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \frac{\partial^2 \phi}{\partial \rho^2} (r - \bar{I}^*(\mathbf{x}, \rho)) \phi(t - \bar{I}(\mathbf{x})).$$

To compute this expression the  $\phi$  function has to be twice differentiable. The previous choice of  $\phi$  defined in section III-A as a third order B-spline is well suited and the derivatives can be developed as follows:

$$\frac{\partial \phi}{\partial \rho} (r - \bar{I}^*(\mathbf{x}, \rho)) = -\frac{\partial \phi}{\partial t} \frac{\partial \bar{I}^*}{\partial \rho} \quad (14)$$

$$\frac{\partial^2 \phi}{\partial \rho^2} (r - \bar{I}^*(\mathbf{x}, \rho)) = \frac{\partial^2 \phi}{\partial t^2} \frac{\partial \bar{I}^*}{\partial \rho} \frac{\partial \bar{I}^*}{\partial \rho} - \frac{\partial \phi}{\partial t} \frac{\partial^2 \bar{I}^*}{\partial \rho^2} \quad (15)$$

The last expression to compute is the derivative of the image intensity with respect to the displacement of the camera. Using both the gradients of the image and the interaction matrix  $\mathbf{L}$  that links the displacement of a point in the image with respect to the camera velocity, the derivatives become:

$$\frac{\partial \bar{I}^*}{\partial \rho} = \nabla \bar{I}^* \mathbf{L} \quad (16)$$

$$\frac{\partial^2 \bar{I}^*}{\partial \rho^2} = \mathbf{L}^\top \nabla^2 \bar{I}^* \mathbf{L} + \nabla \bar{I}^* \mathbf{H}. \quad (17)$$

$\nabla \bar{I}^*$  and  $\nabla^2 \bar{I}^*$  are respectively the gradient and the second order gradient of the current image. Since the only degree of freedom currently considered is the rotation around the vertical axis (the  $y$  axis of the camera), the interaction matrix is then  $\mathbf{L} = [-(1+x^2) \quad -xy]^\top$ , with  $x$  and  $y$  the coordinates of the point  $\mathbf{x}$  ( $\mathbf{x} = (x, y)^\top$ ).  $\mathbf{H}$  is the derivative of the interaction matrix  $\mathbf{L}$  with respect to the rotational velocity. It is defined by [11]:

$$\mathbf{H} = [2x(1+x^2) \quad y(2x^2+1)]^\top.$$

We can notice that previous expressions do not depend on the depth of the scene which means that no 3D reconstruction of the scene is required.

The relation between the real rotation and the computed update is quasi linear. The result of this proposed update will

then cause a quasi exponential decreasing of the error, that is the ideal goal of typical visual servoing tasks [2].

2) *Navigation using visual servoing*: For every acquisition of an image  $I$ , an update  $\delta\rho$  is computed in order to move the camera and increase mutual information between  $I$  and  $I^*$ . Using the model of the vehicle it is possible to go back to the steering angle that will give the estimated update.

Firstly the update of the rotation of the camera is linked to the camera rotational velocity  $\dot{\rho}$  (around the  $y$  axis) by the following equation:

$$\dot{\rho} = \frac{\delta\rho}{\Delta t} \quad (18)$$

where  $\Delta t$  is the processing time (30Hz in our case).

The velocity is directly linked to the steering angle  $\psi$  of the wheels. Using the model of the non-holonomic vehicle used in our experiments (See the car-like model Figure 3) the general steering angle is computed as follows:

$$\psi = \arctan \left( \frac{L \dot{\rho}}{v} \right) = \arctan \left( \frac{L \delta\rho}{v \Delta t} \right) \quad (19)$$

where  $v$  is the translational velocity of the vehicle (along the  $z$  axis) and  $L$  is the distance between the front and rear wheels.

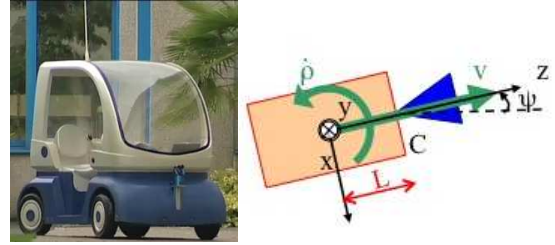


Fig. 3. Considered vehicle and its model.

#### IV. KEY IMAGES SWITCH IN THE VISUAL PATH

To be able to follow the learned trajectory a switching process between the key images of the visual path has to be defined.

A solution could be to consider only the rotation required (given by the parameter update  $\delta\rho$ ) to reach the alignment position. If the computed rotation is smaller than a given threshold it could mean that the vehicle is next to the desired position, then the next key image can be loaded. But such a simple solution will obviously fail when the tracked trajectory is a straight line. The proposed approach is based on this solution coupled with a translation estimation. The key image is updated each time the remaining rotational error is low and that the translation to reach the desired position is small (so that there is no more problem in straight lines).

The rotational error is directly given by the parameter update  $\delta\rho$ . However no estimation is given on the translational error (that is the remaining distance between the current position and the position corresponding to the key image).

The approach is to consider the variation of the mutual information between the key image and the current image

depending on the variation of translational position  $t_z$  of the vehicle. If the variation of MI is null, it means that the maximum of MI is reached and that the robot is at the desired position. If the variation is positive (respectively negative) it means that mutual information is increasing (respectively decreasing) and that the robot is getting close to (respectively moving away from) the desired position.

This variation is simply computed as the derivation of mutual information wrt. the translational velocity  $v_z$  of the camera. The formulation of the problem is similar to the one proposed in section III-B.1 as the difference that the current image is now depending on  $v_z$ . The derivative of the mutual information is now expressed with the interaction matrix corresponding to the translational degree of freedom that is  $L_{t_z} = [x/Z \quad y/Z]^t$  with  $Z$  the depth of each points. Since an accurate estimation of the translation is not needed,  $Z$  is approximated to be constant with  $\bar{Z} = 20$  meters.

To validate the proposed approach, some simulations have been performed using a strongly rough environment. Figure 4 illustrates the performed experiment. The value of the derivative of mutual information is shown depending on the translation between the current and the key position along the  $z$  axis. We can see that the choice of the depth value is not critical (in fact using the previous equations, it can be seen that changing  $Z$  is only modifying the derivative by a scale factor). Considering a strongly non flat scene, mutual information derivative with respect to the translation remains accurate with a null value when the robot reach the desired translation.

Using two given thresholds on both the parameter update and the translation estimation allows to update the key image each time the robot is close to the current desired position.

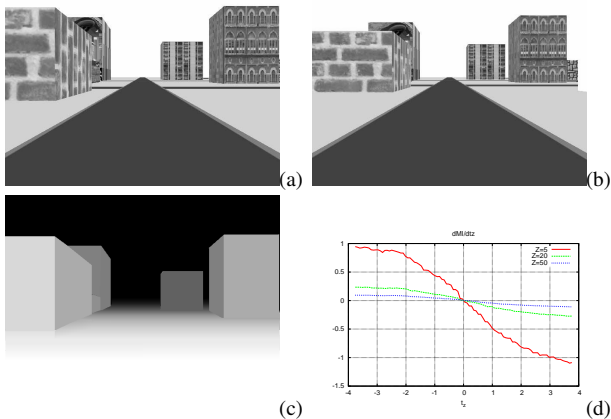


Fig. 4. Translation estimation between the current and desired image. (a) Desired image, (b) acquired image with a 4 meter translation, (c) scene depth and (d) derivative of mutual information with respect to the translation along the  $z$  axis (in meters) with various fixed scene's depth  $Z$ .

## V. EXPERIMENTAL RESULTS

The mutual information navigation scheme has been tested on a non-holonomic vehicle (see Figure 3) in an outdoor environment. The final approach presented in the previous

paragraph has been used. Let us emphasize that the vehicle is equipped with a monocular camera and that no other sensor such as GPS, radar or odometry are considered in these experiments. Furthermore, the 3D structure of the scene remains fully unknown during the learning and navigation steps.

Aerial views of the environment, where the navigation task takes place, are shown in Figure 5 along with the considered trajectory (about 400 meters). As seen on the pictures, the environment is semi-urban with both trees and buildings (with windows acting as repetitive textures). Let us note that the vehicle crosses a covered parking lot (green part of the trajectory in Figure 5) and that the ground is no longer perfectly flat (mainly in the first 100 meters of the trajectory).

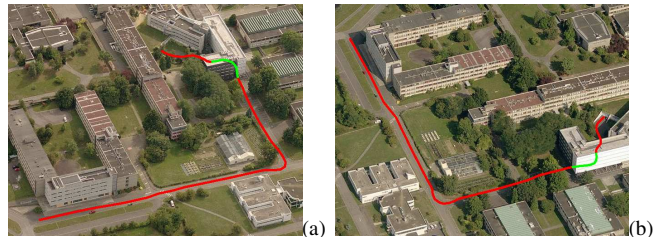


Fig. 5. Outdoor environment with an approximation of the learned trajectory (the red part is the trajectory executed under a covered parking lot). (a) First aerial view, (b) second aerial view.

When learning the path the vehicle is manually driven at a roughly constant velocity. For this experiment we consider 1200 key images (that is around three key images per meter). The navigation task itself is carried out at 0.5 m/s. Images are acquired at 30Hz (nearly 25.000 images are acquired and processed in real-time during this navigation task).

Some pictures of one navigation task are shown in Figure 6. By comparing the current and key images (and the image error on the third row), we can see that the robot is qualitatively (as defined in [15]) following the same path. The navigation task has been tested with both cloudy and sunny weather using the same learned visual path). Since time had passed between the acquisition of the visual path and the navigation task, there have been very large illumination changes between the current and the key images as it is highlighted in Figure 7(a). The task has even been tested with a ground recovered with snow still using the same initial visual path (See Figure 7(b)). Despite those illumination variations the navigation task was still converging. That shows the robustness of the proposed control law to illumination variations and the efficiency of considered mutual-information similarity criterion to perturbation.

Since no obstacle avoidance process is considered, the navigation task has been performed in quiet conditions. Nevertheless several vehicles have overtaken our experimental vehicle and appeared in the camera view. Despite this perturbation, and thanks to the robustness of the similarity criterion, the navigation task has never failed showing the robustness of mutual information to occlusions. One of these moments is represented in Figure 7(c) (the van in the current image was not present in the key image). The complete video

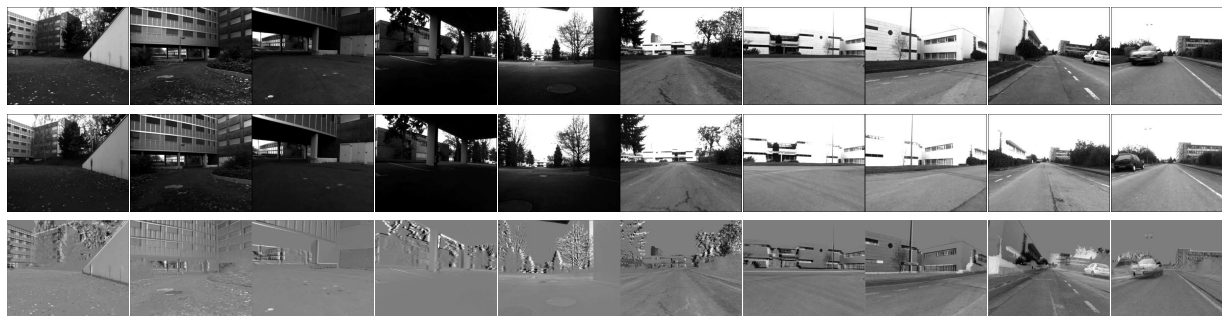


Fig. 6. Outdoor navigation experiment. First row: current image acquired by the vehicle, second row: desired image and third row: difference between the current and desired image. The first top column and the last bottom column show respectively the first images and the last images used in the navigation task.

is provided with submission.

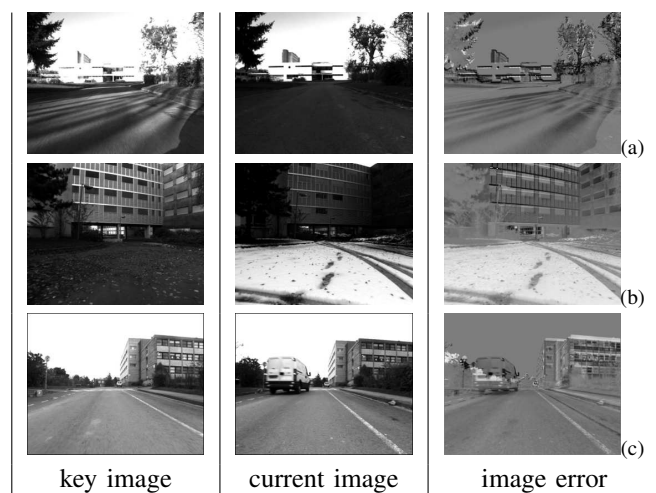


Fig. 7. Mutual information robustness. (a) robustness to illumination variations, (b) illumination variations and snow on the ground, (c): robustness to occlusions. First column: desired image, second column: current image and third column: difference of the current and desired images.

## VI. CONCLUSION

In this paper, we have presented a new way to achieve image based navigation task. We show that our vehicle is able to track a previously learned trajectory using only the information provided by a monocular camera. The navigation task can be achieved despite important variation in the lighting condition and possible perturbations. This can be achieved thanks to various elements related to the use of mutual information which is classical in medical image registration but new in vision-based control:

- Our approach does not rely on features extracted from the image. Therefore we do not need to track or match features (eg, keypoint) which has proved to be a difficult and not always reliable process. Furthermore no 3D information related to the scene structure is required.
- To avoid this tracking and matching processes, the vision-based control law of the non-holonomic vehicle is directly linked to the optimization of a similarity criterion based on the information shared by two images.

- Considering information contained in the images and not features extracted from the image or the image intensities induces a natural robustness to perturbation that is essential in our navigation context.

## REFERENCES

- [1] S. Baker, I. Matthews. Equivalence and efficiency of image alignment algorithms. *IEEE CVPR'01*, pp. 1090 – 1097, Dec. 2001.
- [2] F. Chaumette, S. Hutchinson. Visual servoing and visual tracking. *Handbook of Robotics*, Springer, 2008.
- [3] Z. Chen, S.T. Birchfield. Qualitative vision-based path following. *IEEE Transactions on Robotics*, 25(3):749–754, 2009.
- [4] L.A. Clemente *et al.* Mapping large loops with a single hand-held camera. *RSS'07*, Atlanta, Jun. 2007.
- [5] C. Collewet, E. Marchand. Photometric Visual Servoing. *IEEE T. on Robotics*, 2011.
- [6] J. Courbon, Y. Mezouar, P. Martinet. Autonomous navigation of vehicles from a visual memory using a generic camera model. *IEEE Trans. on Intel. Transportation Systems*, 10(3):392–402, 2009.
- [7] A. Dame, E. Marchand. Entropy based visual servoing. *IEEE ICRA'09*, pp. 707–713, Kobe, May 2009.
- [8] A. Diosi, *et al.* Outdoor visual path following experiments. *IEEE/RSJ IROS'07*, pp. 4265–4270, San Diego, Oct. 2007.
- [9] N.D.H. Dowson, R. Bowden. A unifying framework for mutual information methods for use in non-linear optimisation. *ECCV'06*, pp. 365–378, Jun. 2006.
- [10] U. Frese, L. Schröder. Closing a million-landmarks loop. *IEEE/RSJ IROS'06*, pp. 5032–5039, Beijing, Oct. 2006.
- [11] J.-T. Lapresté, Y. Mezouar. A Hessian approach to visual servoing. *IEEE/RSJ IROS'04*, pp. 998–1003, Sendai, Oct. 2004.
- [12] F. Maes, *et al.* Multimodality image registration by maximization of mutual information. *IEEE T. on Med. Imaging*, 16(2):187–198, 1997.
- [13] G. Panin, A. Knoll. Mutual information-based 3d object tracking. *IJCV*, 78(1):107–118, 2008.
- [14] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE T. on Medical Imaging*, 22(8):986–1004, Aug. 2003.
- [15] A. Remazeilles, F. Chaumette. Image-based robot navigation from an image memory. *Robotics and Autonomous Systems*, 55(4):345–356, April 2007.
- [16] E. Royer, *et al.* Monocular vision for mobile robot localization and autonomous navigation. *IJCV*, 74(3):237–260, 2007.
- [17] S. Se, *et al.* Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *IJRR*, 21(8):735, 2002.
- [18] S. Segvic, *et al.* A mapping and localization framework for scalable appearance-based navigation. *CVIU*, 113(2):172–187, February 2009.
- [19] C.E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.
- [20] P. Thévenaz, M. Unser. Optimization of Mutual Information for Multiresolution Image Registration. *IEEE T. on IP*, 9(12), 2000.
- [21] P. Viola, W. Wells. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, 1997.
- [22] Y.-M. Zhu. Mutual information-based registration of temporal and stereo retinal images using constrained optimization. *Comput. Methods Prog. Biomed.*, 86(3):210–215, 2007.