Robotica (2017) © Cambridge University Press 2017
doi:

Ground appearance-based monocular visual odometry for planar motion estimation

Richard Béarée† * and Franck Hernoux†

†*LSIS CNRS 7296, Arts et Métiers campus Lille, 59046 Lille FRANCE*

(Accepted MONTH DAY, YEAR. First published online: MONTH DAY, YEAR)

SUMMARY
In this paper we propose a practical method for computing the planar motion estimation of vehicles, specifically for indoor relative self-localization of industrial mobile platforms. The motion information is extracted from a downward-looking monocular camera fixed to the vehicle without any constraint on the vehicle steering geometry. Based on normalized cross-correlation metric, multi-template matching for consecutive images gives the camera pose, i.e. the position and the heading orientation corresponding to the rigid planar motion of the vehicle. This algorithm takes advantages of the robustness of the cross-correlation metric to small angular changes for a predefined template size. In order to accommodate for higher speed systems, a dynamic selection of a two-template configuration is proposed, which increases the reachable angular and linear velocity in all directions. Finally, the accuracy performances of the proposed visual odometry system are experimentally tested on indoor concrete surface using omnidirectional industrial platform.

KEYWORDS: Localization; Visual Odometry; Computer Vision; Wheeled Robots; Autonomous Vehicles.

1. Introduction

In the field of industrial and service robotics, use of Automatic Guided Vehicles or mobile platforms for indoor applications is becoming more prevalent. Mobile platforms, which can be associated to one or more collaborative robotic arms, open the door to new and more flexible industrial applications in warehouses or assembly lines. The first capability a mobile robot must have for autonomous operation is to accurately solve the self-localization problem. For years, mechanical and optical wheel encoder systems have been widely used to solve the localization problem for autonomous vehicles. The main advantages of these odometry systems are their simplicity and their robustness to external influences. However, wheel odometry relies on the assumptions that there is no ground slippage phenomenon and that the kinematic model of the vehicle is perfectly known and stationary. For flexibility and compactness reasons, recent mobile robots take advantage of the omnidirectional motion possibility offered by Omni-wheels or Mecanum wheels to move instantaneously in any direction.[1] For this kind of vehicle, depicted in Fig. 1, the free rollers elements induce slip motion and the use of wheel encoder is inherently inadequate.[2] This problem can be considered easy to solve as accurate sensors such as Inertial Measurement Unit or laser-based distance measurement system (LIDAR) associated to elegant and efficient algorithms such as Simultaneous Localization And Mapping have been successfully implemented in field applications.[3,4] However, previous luxury solution could be a serious impediment to widen the use of mobile platforms for small and medium-sized enterprises. Therefore, there is a growing and imperative need for low-cost localization solutions, which also can be used for

* Corresponding Author. E-mail:richard.bearee@ensam.eu

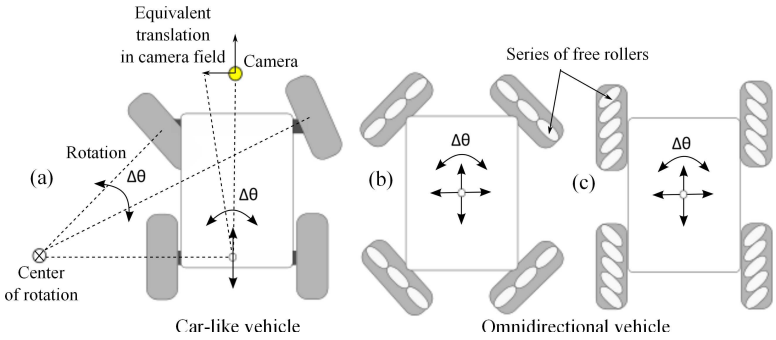

Fig. 1. (a) Ackermann steering model. (b) Omni wheels. (c) Mecanum wheels.

retrofitting of existing AGVs. The specification associated to stringent constraints ensuring the industrial viability need can be detailed as follows: (a) the unavailability of GPS for indoor applications; (b) the easiness of implementation without warehouse facility modifications. The main reason is economical. Other significant benefit is the possibility to move the system to different buildings, while reducing the technical installation difficulties; (c) the solution should be independent of the vehicle dimensions or steering models (holonomic or car-like); (d) the presence of humans working around the system has to be considered (not only for security reason, but for visual occlusion problems during collaboration too); (e) the overall cost of the system should be limited.

Based on the previous constraints, this paper proposes a monocular vision-based system adapted to the odometry for ground vehicles, independent of their steering characteristics (car-like vehicle, holonomic omnidirectional vehicle). The camera is directed towards the ground and takes advantages of fast multi-template correlation to estimate the 2D vehicle motion, i.e. the camera position and orientation resulting from the relative planar ground motion. The developed visual odometry solution include, but is not limited to indoor applications, and is totally insensitive to slippage phenomenon. The proposed substitute for mechanical odometers can be used as an input in a multi-sensors algorithm to provide improved localization information. However, it is mainly intended to be used in warehouse for self-localization between known workstations. The loop closing mechanism to reset the location can be done with, for example, local passive RFID tags located in each workstation. The relation between the proposed solution and the list of specifications will be made clear in the next section devoted to the related works.

The remainder of this paper is divided into four sections. Section 2 presents the state of the art of the practical visual odometry and positions our works. Section 3 describes the proposed multi-template correlation-based visual odometry. Section 4 proposes some adaptations of the algorithm to increase the maximum reachable velocity for a given image resolution. Finally, section 5 evaluates the proposed algorithm on a real industrial mobile platform.

2. Related work

Stereo Visual Odometry is an historical and efficient approach to derive the relative motion of ground robot.[5,6] In [7], authors propose an indirect method estimating the 6DoF egomotion of a forward-looking stereo camera. Estimation is based on the trifocal geometry between image triples and experiments conducted over long path exhibit a very goog accuracy in the range of 1.5 %. Stereo cameras help provide scale and some constraints to assist recovery of motion. The counterpart is that the cost in components, interfacing, synchronization, and computing are higher for stereo cameras compared to a monocular camera.[8] Over the past decade, many monocular approaches

have been developed. In [9], authors describe a visual odometry method, which uses an omnidirectional camera mounted on the roof of the vehicle. The proposed robust hybrid method for structure from motion, combines the performances of feature-based approach for the translation estimation of the ground with the robustness and higher resolution of appearance-based method for the heading orientation. In [10], an efficient and low-cost solution for the 2D absolute localization of outdoor ground vehicle is presented. A fast whole image alignment approach is developed using the image sequence from a standard rear parking camera as input, but assuming a constant velocity model for the dynamics of the vehicle. Some works restrict the degrees of freedom of the odometer by using a specific nonholonomic system model (e.g. Ackerman steering model). This widely used assumption for nonholonomic robots reduces the camera motion possibility and transforms the vehicle rotation around rear axle into lateral translation in the camera frame. In [8], authors use a forward-looking monocular approach, with a nonholonomic steering model assumption to separately recover rotation and translation. The visual information is fused with wheel odometry to provide robust estimation of the 3D pose of a mobile platform in undulating terrain. Experiments indicate accuracy comparable to that from state-of-the-art stereo systems, i.e. around 1 %. One notes that the reported accuracy for the visual odometry alone is between 1.5 % and 4 %. Among the others methods for fast planar motion estimation, low-cost solutions using optical mouse sensor can be cited.[11,12] To recover the position and orientation of the system at least two optical mice should be used.[13] In [14], Author uses a redundant number of optical mice (≥ 2) and assuming restrictions on the installation of optical mice at the bottom of a mobile robot, the optimal sensor placement for velocity estimation is discussed. However, the main drawback of optical mouse odometry lies in the small working focal range of such a sensor. The localization accuracy is sensitive to small variation of the ground distance (e.g. motion-induced vibration). In [15] the problem is partially overcome by using several mouse sensors each attached to a lens of a different focal length. Linear speed is estimated using the optical sensor and the heading direction comes from the attitude determined by integrating IMU.

Among the huge number of works and algorithms regarding Visual Odometry for ground robot, a few works have been done considering a single ground-facing camera approach. The solution to the problem of planar motion estimation (one translation and one rotation) for a monocular ground-facing camera can be coarsely solved using three methods. The first method is a brute force algorithm using appearance-based approach. This direct method, in contrast to features-based approach,[16] is here based on image intensity correlation algorithm using cross-correlation metrics in spatial domain, or its frequency-domain counterpart with the Fourier transform-based phase-correlation method.[17] This method estimates the vehicle motion by analyzing the shift of an image region, called a template, in two consecutive images. For two consecutive images, a selected template is iteratively rotated by a predefined angular increment and the best cross-correlation value gives the estimated rotation and translation of the camera. This algorithm is used for example in [18] for the ground visual odometry of omnidirectional vehicles. Experimental results, obtained by off-line processing, confirm the sufficient accuracy of the method, as compared to wheel odometry. However, the significant number of cross-correlation operations, required to reach a reasonable angular resolution, makes this brute force methodology not conceivable for real-time applications. The second method consists in using the phase-correlation algorithm in a two-stage approach to estimate the rotation, and then the translation. The method exploits the Fourier shift properties and the Fourier-Melin Invariant, which respectively state that the magnitude of the Fourier transform is invariant to translations and the log-polar space maps to rotation and scale.[19,20] In spite of the use of the fast Fourier transform and other optimizations, this method used in [21] for a Rover's localization problem, is still not real-time compliant considering a suitable angular resolution. For instance, for an expected angular resolution of only 0.2 degree, the template size has to be at least of 573 pixels, and the computing time is in the

order of 0.1 seconds.[22] The third approach takes account implicitly of the steering model of the mobile robot as previously mentioned for the forward-facing approach. Thus, only longitudinal and lateral translation shifts in the camera field has to be estimated (see Fig. 1(a)). In [23], an Optical Flow odometry method is analyzed. The method fuses wheel encoder odometry and feature-based OF odometry obtained from a monocular camera looking at the ground. Authors report an accuracy problem limiting significantly the vehicle velocity, which is mainly due to the reduced size of the searching area for OF. In the same context, the works in [24], investigating OF odometry obtained from the well-established pyramidal Lucas-Kanade algorithm,[25,26] indicates the same accuracy limitation for a single downward-pointing camera. The performances of the feature-based approach will be intrinsically dependent of the quality and number of features, which can be very restricted considering only ground surfaces, such as low texture asphalt or concrete floor.[15] In [27], authors introduced the concept of correlation-based visual odometry for car-like vehicle, using a downward-looking camera on the front of the vehicle, perpendicular to the ground. The real-time performances and the accuracy of the proposed method are improved in the recent work [28], by adding a fast template selection method based on a template quality measure, and a motion predictor filter, which restricted the search area for the template matching.

Finally, the correspondence between the state-of-the-art in mobile platform localization and the list of practical specifications given in introduction can be briefly summarized as follows. For the versatility of the odometry solution, the method has to be independent of the system motion model. For ease of implementation, ease of tuning and low operating costs, the monocular solution seems prevalent. Hence, to avoid visual occlusion problem, unavoidable in the context of an industrial platform working alongside humans on the factory floor, the downward-looking approach is investigated in this study. Moreover, shadows phenomenon which highly influenced the performance of any vision-based approach can be significantly reduced using the ground-facing camera approach combined with appropriated light sources. As a final point, the reported robustness property of the correlation-based method for low-textured surfaces as compared to OF is a key argument. In the remainder of this paper, we propose a simple and pragmatic visual odometry system using a downward-looking monocular camera. The algorithm is based on the same practical cross-correlation method used in [27], but extend the results to the omnidirectional vehicle case (rotation estimation), by exploiting multi-template correlation in each successive image to quickly estimate the translation and rotation of the camera frame.

3. Rigid planar motion estimation using multi-template matching approach

3.1. Planar motion odometry

In our applications, we suppose that the vehicle or mobile robot moves in planar ground and the camera distance from the ground h is constant on average of a travel, thus the scale difference between successive frames can be reasonably neglected. Figure 2 shows the parameters used to describe the planar motion of the camera frame. The parameters which are used to uniquely identify the motion of the camera center point at time k/F_s (F_s being the acquisition frequency of the camera) are the position vector $\mathbf{p}_k$ and the angle θ_k. Then, two consecutive frames I_{k-1} and I_k will be only related by a combination of one translation $\delta\mathbf{p}_k$ and one rotation $\delta\theta_k$. The current camera frame location and orientation will be given by

$$\begin{aligned} \mathbf{p}_k &= \mathbf{p}_{k-1} + \mathbf{R}(\,\delta\theta_k) \cdot \delta\mathbf{p}_k \\ \theta_k &= \theta_{k-1} + \delta\theta_k \end{aligned}, \tag{1}$$

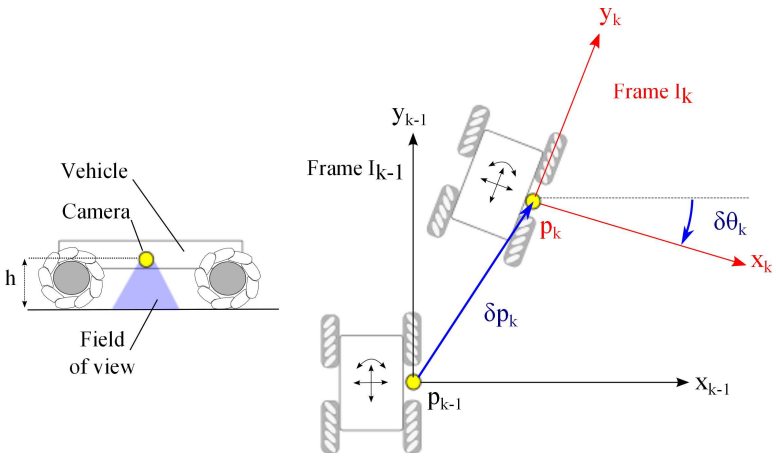

Fig. 2. Parameters of the visual planar odometry. The camera location is an example of configuration for omnidirectional vehicle.

with $\mathbf{R}(\delta\theta_k)$ the rotation matrix

$$\mathbf{R}(\delta\theta_k) = \begin{pmatrix} \cos(\delta\theta_k) & \sin(\delta\theta_k) \\ -\sin(\delta\theta_k) & \cos(\delta\theta_k) \end{pmatrix}. \tag{2}$$

The conversion of the displacement $\mathbf{p}_k$ from pixels in the camera field to real distance $\mathbf{d}_k$ in meters can finally be expressed as

$$\mathbf{d}_k = C_h \cdot \mathbf{p}_k, \tag{3}$$

with $C_h = h/(f \cdot p_S)$ the basic length unit for the odometry system, p_S being the pixel size, f the camera focal-length and h the camera distance to the ground. One note that the previous intrinsic camera parameters can effectively be retrieved during the calibration stage of the camera. However, after integration of the radial distortion correction, the easiest way for practical scaling of the system consists in identifying directly the parameter C_h by travelling a known distance with the vehicle.

The extraction of the rotational motion $\delta\theta_k$ for succesive frames is central to our vehicle navigation system. Consequently, before detailing the proposed visual odometry, the rotation influence on the template matching algorithm is reviewed.

3.2. Fast template matching and rotational motion influence

3.2.1. NCC template matching. Normalized Cross Correlation (NCC) is a well-established matching algorithm,[29–32] which represents the intensity similarity between a reference image and another image or sub-image, called template. Let I_{k-1} and I_k be two consecutive images of size $M_c \times M_r$ and T_{k-1} a choseen template of size $N_c \times N_r$ extracted from I_{k-1}. The NCC value Γ at each point (u, v) for I_k and the template T_{k-1}, which has been shifted by u pixels in x direction and by v pixels in y direction, can be written

$$\Gamma(u,v) = \frac{\sum_{x,y} I_k(x,y) \cdot T_{k-1}(x-u, y-v)}{\sqrt{\sum_{x,y} I_k^2(x,y) \cdot \sum_{x,y} T_{k-1}^2(x-u, y-v)}}. \tag{4}$$

The estimated position $\mathbf{t}_{est,k} = (u_{est,k}, v_{est,k})$ of the template T_{k-1} in the image I_k will be given by

$$\mathbf{t}_{est,k} = \arg\max_{u,v} (\Gamma(u,v)). \tag{5}$$

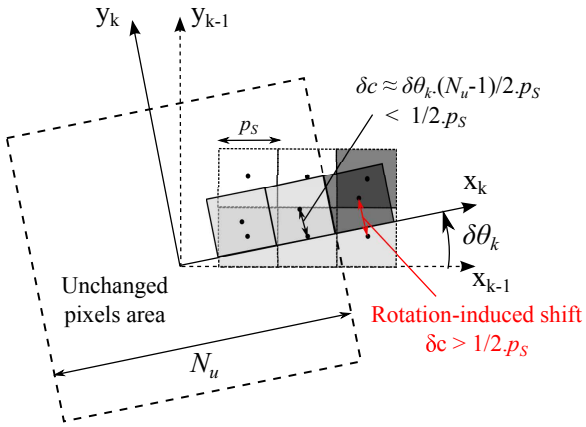

Fig. 3. Illustration of the rotation-induced shift.

If the template matching is successful, it should be noted that the estimated position $\mathbf{t}_{est,k}$ is equal to the position of the template T_{k-1}, noted $\mathbf{t}_{k-1}$, expressed in the current frame k of I_k, i.e. $\mathbf{t}_{est,k} = [\mathbf{t}_{k-1}]_k$. Hence, the template shift, noted $\delta\mathbf{T}_k$, can be finally calculated as

$$\delta\mathbf{T}_k = \mathbf{t}_{est,k} - \mathbf{t}_k = [\mathbf{t}_{k-1}]_k - \mathbf{t}_k. \tag{6}$$

To reduce the brightness variation influence, one commonly uses the normalized version of the cross correlation coefficient, where the mean intensity values are subtracted from both image inputs, i.e. $I(x,y)$ and $T(x,y)$ are respectively replaced by $I'(x,y) = I(x,y) - \overline{I}$ and $T'(x,y) = T(x,y) - \overline{T}$ with

$$\overline{I} = \frac{1}{M_c M_r} \sum_{x,y} I(x,y),\ \overline{T} = \frac{1}{N_c N_r} \sum_{x,y} T(x,y). \tag{7}$$

This version of the algorithm being more robust to noise and brightness fluctuations,[33] we will use it in the remainder of this paper.

3.2.2. Rotation influence on ground images cross-correlation. By definition, the previous NCC metric approach is efficient for translation shifts measure and it is only used for non-rotational environment. Indeed, rotation between image and template frames modifies the pixel's location, which may rapidly cause inaccurate template matching. Assuming that the pixels intensity values after rotation can be obtained by nearest neighbor interpolation, the number of pixels that are not influenced by rotation (i.e. unshifted) can be approximated using the parameterization of Fig. 3. A pixel intensity value will be unchanged as long as its center point remains at a distance δ_c lower than almost $1/2 \cdot p_S$ from its unrotated counterpart, with p_S the pixel size. Assuming small rotation $\delta\theta_k$ performed around the center of a square template of size N, the distance δ_c for a pixel center initially located at x_0 along the x_{k-1} axis can be calculated as: $\delta_c = x_0.\sin(\delta\theta_k) \approx x_0.\delta\theta_k$. Considering now N an even number of pixels, the size N_u of the unchanged pixel area verifies the following inequality: $(N_u - 1)/2 \cdot p_S \cdot \delta\theta_k < p_S/2$. This leads to an approximation of the unchanged pixels number (with $\lfloor\alpha\rfloor$ the integer part of α):

$$N_u = \left\lfloor \frac{1}{\delta\theta_k} \right\rfloor + 1. \tag{8}$$

In the following, the influence of rotation on the NCC template matching performances is experimentally analyzed for the case of low-textured ground images. To clearly show the influence of rotation on the ground image template matching, we have used images

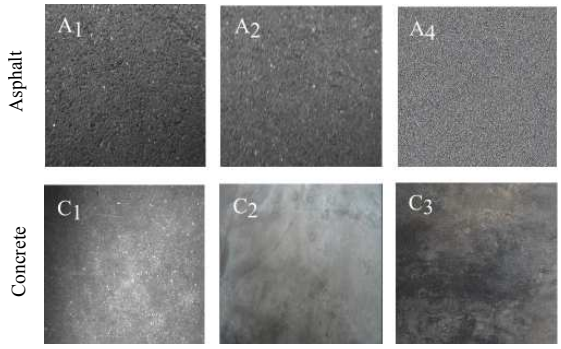

Fig. 4. Samples of asphalt and concrete ground images (different surfaces and different levels of illumination).

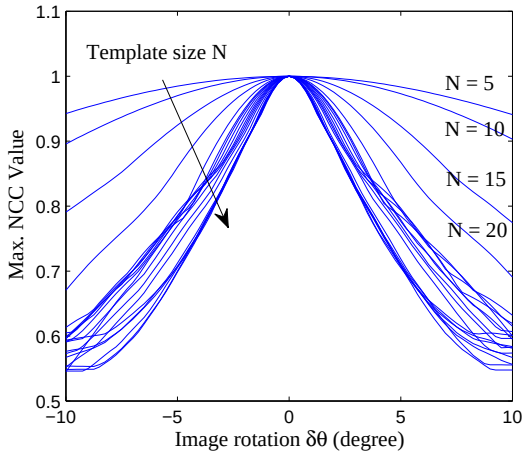

Fig. 5. Example of the rotation influence on the maximum NCC value. The template mask size N is between 5 and 100 pixels and the surface is the low-textured concrete floor noted C_1 in Fig. 4).

of concrete and asphalt surfaces, with different brightness conditions (see Fig. 4). For each ground type, the experiments were performed using real rotated images to avoid interpolation effect of mathematically rotated images. A CCD camera was mounted and aligned on a high precision rotary table (angular resolution below 0.0035 degree) and the camera axis was rotated incrementally by step of 0.5 degree. Templates with varying size N located in the center of the rotated images were defined. Finally, template matching using NCC metric was done for each template and a result other than zero means a rotation-induced shift (false translation shift). Fig. 5 shows an example of evolution of the maximum NCC peak value according to the rotation value and for different template sizes N between 5 and 100 pixels. Clearly, the rotation decreases the maximum NCC value. This decrease, which is of approximately 4% per degree for N higher than 30 pixels, is mainly due to the equivalent rotation filtering effect.

Fig. 6 shows the relation between the template size N and the maximum image rotation free from false shift for a set of concrete, asphalt and synthetic images. The synthetic images, which are only used here for comparison, are composed of random intensity pixels values generated by Normal distribution. In fact, the intensity distribution of low-textured ground images can be reasonably approximated as a normal distribution.[28] The thin dashed lines represent an estimated percentage of unrotated pixels according to the rotation angle and the template size as given by the relation (8). Lastly, all the statistical results for the template sizes of 40 and 80 pixels are given in Table I. Fig. 6 and Table I show that for limited angles and limited template sizes, the cross-correlation of uniform and untextured images can be considered as nearly rotation-invariant. The maximum

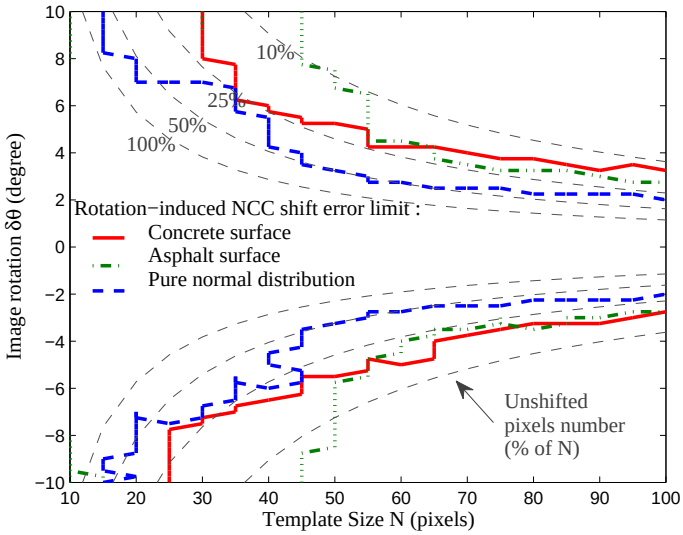

Fig. 6. Limit of the image rotation that induces no cross-correlation translation shift according to the template mask size. Concrete, asphalt and synthetic image (Normal distribution) are compared. The thin dashed lines represent an estimation of the percentage of the template pixels, which are not influenced by rotation.

Table I . Maximum rotation value for NCC template matching free from false shift.

	$\max\|\delta\theta_k\|$: mean/median/std	
	$N = 40$	$N = 80$
Asphalt	13.3°/12.1°/0.7°	3.2°/3.1°/0.4°
Concrete	6.2°/6.6°/0.6°	3.1°/3°/0.3°
Synthetic image	4.5°/4.4°/0.3°	2.2°/2.2°/0.2°

rotation value depends on the surface type. For the synthetic images, the rotation does not induce translation shift as long as almost half of the template pixels are unrotated (see the thin dashed line with the 50% tag in Fig. 6). For real concrete ground images, this limit is closer to 25% of unrotated pixels. Finally, templates matching of asphalt images are even more robust to rotation. Qualitatively speaking, the more the image is textured (sharpness and entropy metrics), the bigger the rotation-invariant area. In this study, we keep in mind that for concrete and asphalt floor, which are common surfaces in factory and warehouse, the NCC matching of a template of almost 40 pixels size can be considered as insensitive to a template rotation less than ±4 degrees.

The previous result is the fundamental hypothesis of the proposed visual odometry system. The template size will be set to 40×40 pixels and the maximum rotation for consecutive images is limited to ±4 degrees (see. Fig. 6). The limitation of the rotation angle needs to be put into perspective. For a camera acquisition frequency F_s of 30 Hz or 60 Hz, the maximum value of the instantaneous angular velocity in the camera field could be respectively of 120 degree/s (2.1 rad/s) or 240 degree/s (4.2 rad/s), which is beyond the specifications of classical industrial or service AGV. For higher speed systems, as will be shown later, this limitation can be circumvented using pre-rotation of the image I_{k-1}. Finally, concerning the chosen template size, it has proven reliable on all the ground surfaces we have tested. The template size being set, the cross-correlation processing time will only depends on the size of the research area. For a full image of 640×480, the processing time is around 70 ms, which is not real-time. But, as we will see later, the search area can be significantly reduced according to the apparent motion direction in

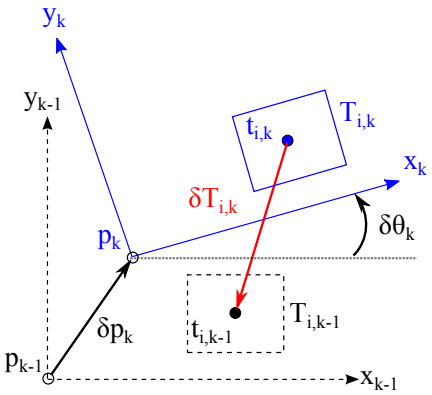

Fig. 7. Parameters of the template matching.

the camera frame. The restriction of the search area in the full image reduces drastically the processing time, which becomes largely real-time. For instance, the computing time for a search area of 200×200 pixels will be less than 5 ms.

3.3. Principle of the proposed visual odometry using multi-template matching

Exploiting the previous template matching, the two parameters of the template translation $\delta\mathbf{T}_k$ between consecutive images can be calculated. Hence, in order to estimate the three parameters $\delta\mathbf{p}_k = [\delta p_{x,k}\ \delta p_{y,k}]^T$ and $\delta\theta_k$ of the vehicle relative motion, at least two template matchings have to be done at different image locations. Although the proposed method employs preferentially the two-template configuration for computing performance, the general multi-template configuration is detailed in the following paragraph.

For each image frame k, we define n template masks $T_{i,k}$, $i = 1, \cdots, n$, with their respective center locations in image I_k noted $\mathbf{t}_{i,k} = [t_{ix,k}\ t_{iy,k}]^T$. The definition of the parameters for the i^{th} template are given in Fig. 7. Assuming that the rotation is small between consecutive images, as described in the previous part, the NCC matching of each template gives the translation vector $\delta\mathbf{T}_{i,k} = [\delta T_{ix,k}\ \delta T_{iy,k}]^T$ (see relation (6)). Recalling that this vector results from the combination of the rigid transformation $(\delta\mathbf{p}_k, \delta\theta_k)$ applied to the previous location of each template $\mathbf{t}_{i,k-1}$, the following relation can be written

$$[\mathbf{t}_{i,k-1}]_k = \mathbf{t}_{i,k} + \delta\mathbf{T}_{i,k} = \mathbf{R}(-\delta\theta_k) \cdot \mathbf{t}_{i,k} - \delta\mathbf{p}_k. \tag{9}$$

For small angle $\delta\theta_k$, $\mathbf{R}(-\delta\theta_k)$ can be approximated as

$$\mathbf{R}(-\delta\theta_k) \approx \begin{bmatrix} 1 & \delta\theta_k \\ -\delta\theta_k & 1 \end{bmatrix}, \tag{10}$$

and Eq. (9) can be rewritten as

$$\mathbf{A} \cdot \delta\mathbf{x}_k = \delta\mathbf{T}_k, \tag{11}$$

with the notations $\delta\mathbf{x}_k = [\delta p_{x,k}\ \delta p_{y,k}\ \delta\theta_k]^T \in \mathbb{R}^{3\times 1}$, $\delta\mathbf{T}_k = [\delta T_{1,k}\ \cdots\ \delta T_{n,k}]^T \in \mathbb{R}^{2n\times 1}$ and $\mathbf{A} = [\mathbf{A}_1\ \cdots\ \mathbf{A}_n]^T \in \mathbb{R}^{2n\times 3}$, with

$$\mathbf{A}_i = \begin{bmatrix} -1 & 0 & -t_{ix,k} \\ 0 & -1 & -t_{iy,k} \end{bmatrix}. \tag{12}$$

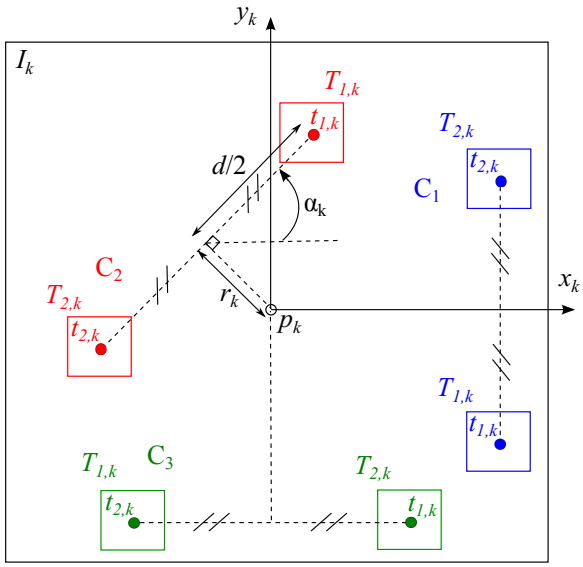

Fig. 8. Parameters for the two-template locations. The three main configurations are shown: C_1 along the x-axis ($\alpha_k = \{\pi/2, 3\pi/2\}$), C_2 along the image diagonal ($\alpha_k = \{pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$), and C_3 along the y-axis ($\alpha_k = \{0, 2\pi\}$).

Since $n \geq 2$, the system (11) is overdetermined and the least square solution can be calculated using the Moore-Penrose pseudoinverse of $\mathbf{A}$

$$\delta \mathbf{x}_k = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \cdot \delta \mathbf{T}_k. \tag{13}$$

Eq. (13) gives a general solution to the problem of the least square estimation of the camera rigid body motion for the multi-template case. For particular configuration of the templates, the analitycal solution of (13) can be simplified. The interested reader is referred to similar approaches detailled in [12,14], where a polygonal array of optical mice is used for the velocity estimation of mobile robots. As already mentioned, in order to reduce the odometry computing cost of the proposed solution, the number of templates is preferentially limited to $n = 2$. Fig. 8 shows the three parameters that determine the two-template configuration. d is the distance in pixels between the two templates center points, $\alpha_k = j\pi/4$ ($j \in \mathbb{N}$) is an angle that defines the orientation of the configuration and r_k is the distance in pixels between the image center point and the two-template axis. The template $i = 1, 2$, position vector in the image frame k can be expressed as

$$\mathbf{t}_{i,k} = \begin{bmatrix} t_{ix,k} \\ t_{iy,k} \end{bmatrix} = \begin{bmatrix} \frac{d}{2}\cos(\alpha_k + (i-1)\,\pi) + r\sin(\alpha_k) \\ \frac{d}{2}\sin(\alpha_k + (i-1)\,\pi) + r\cos(\alpha_k) \end{bmatrix}. \tag{14}$$

Finally, the odometry solution of Eq. (13) for the two-template configuration can be given as follows

$$\delta p_{x,k} = -\frac{\Sigma_x}{2} + r_k \cdot \sin(2\alpha_k)\frac{\Delta_x}{2d} - r_k \cdot \cos^2(\alpha_k)\frac{\Delta_y}{d}, \tag{15}$$

$$\delta p_{y,k} = -\frac{\Sigma_y}{2} + r_k \cdot \sin(2\alpha_k)\frac{\Delta_y}{2d} - r_k \cdot \sin^2(\alpha_k)\frac{\Delta_x}{d}, \tag{16}$$

$$\delta \theta_k = \sin(\alpha_k)\frac{\Delta_x}{d} - \cos(\alpha_k)\frac{\Delta_y}{d}, \tag{17}$$

with the notations $\Sigma_x = \delta T_{1x,k} + \delta T_{2x,k}$, $\Sigma_y = \delta T_{1y,k} + \delta T_{2y,k}$, $\Delta_x = \delta T_{1x,k} - \delta T_{2x,k}$ and $\Delta_y = \delta T_{1y,k} - \delta T_{2y,k}$.

Hence, the proposed two-template matching procedure make available the parameters of the camera rigid motion $\delta\mathbf{p}_k$ and $\delta\theta_k$ using (15), (16) and (17). Some remarks can be made. First, when the two templates are symmetrically placed about the image center point, i.e. for $r_k = 0$, $\delta\mathbf{p}_k$ is determined as the average of the two templates translations. Second, the rotation angle $\delta\theta_k$ is inversely proportional to the distance d between the two templates. It follows therefore that the value of d determines the minimum rotation increment and, as such, also the angular accuracy. For experiments, the basic angular unit has been set to the value of 0.25 degree (0.0044 rad). Then, the corresponding distance d between the center points of the templates has to be set higher than 230 pixels ($\approx 1/0.0044$). In the following d is fixed to 240 pixels.

The proposed solution, as any odometry algorithm, is subject to cumulative errors. Moreover, the template matching operation can introduce false estimation or fail for a very low-textured image area, which can induce shot or impulse noise. One way to easily remove such outliers, consist in applying a low-pass filter or a median filter to the template translation measure. In this work median filter with a limited windows size (3 to 5 samples) was preferred, because it preserved better the shape of the signal.[34] As previously mentioned, the template size N being set to 40 pixels, the computing time will depend on the size of the search area in the full image. Hence, we voluntarily restrict the search area to 240 pixels for each template, which conducts to a computing time of almost $5 \sim 7$ ms per template, i.e. a computing frequency higher than 60 Hz for two template-matching processing. With this fixed research area, the maximum template shift in all directions for consecutive images is of 100 pixels, and any of the two-template configuration symmetrically placed about the image center point can be used interchangeably. Using the relation (3) the maximum linear velocity can be calculated. For instance, with the CCD camera parameters used in this paper and located approximatively 600 mm above ground level, the maximum reachable velocity for the predefined search area would be of 12 km/h (3.4 m/s) at an acquisition frequency F_s of 30 Hz. The maximum linear velocity being linearly dependent on F_s and inversely proportional to the focal length f of the camera, these two main parameters has to be chosen according to the desired maximum value of the linear velocity. However, as detailed in the following part, the maximum measurable linear and rotational velocities, for a given set of camera parameters, can be increased to accommodate for vehicles with higher speed, while not sacrificing the computing time performances.

4. Some considerations on the adaptation to higher speed systems

4.1. Motion estimation and Two-template configuration selection to increase linear speed

To increase the maximum allowable velocity for the vehicle, the two-template configuration can be dynamically selected to increase the templates shift possibility in the motion direction. For a given distance d of 240 pixels between the two templates, the configuration is defined by the two remaining parameters r_k and α_k (see. Fig. 8). In Fig. 9, the available space for the three configurations described in Fig. 8 are superimposed. The width and length of the free space for each configuration are $d_1 = (N - M)/2 - d/(2\sqrt{2})$, $d_2 = N - M$ and $d_3 = (N - M - d)/2$, with M and N respectively the full image size and the template size. Considering $M = 480$ pixels, $N = 40$ pixels, Fig. 10 presents the maximum displacement according to the choice of the two-template configuration and the direction of the apparent vehicle motion in the camera frame given by the current motion angle β_k

$$\beta_k = \angle(\delta\mathbf{p}_k)\,. \tag{18}$$

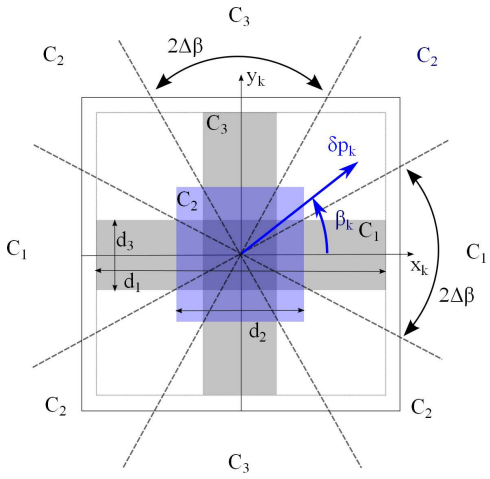

Fig. 9. Apparent motion parameters and free space representation according to the two-template configuration.

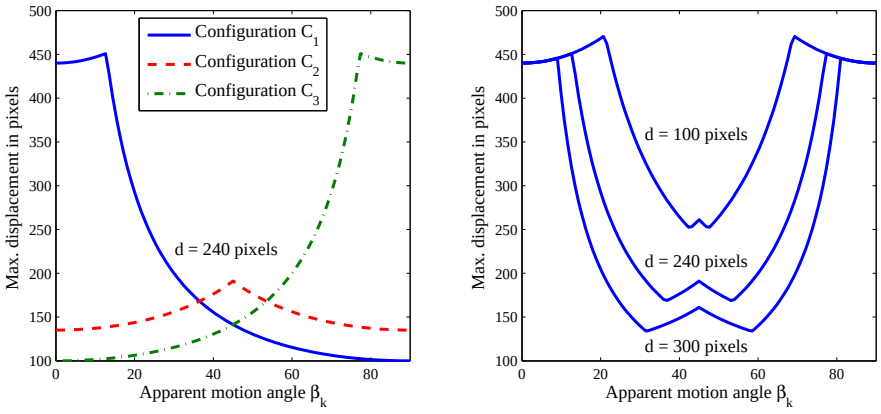

Fig. 10. Maximum template displacement according to the choice of the two-template configuration and the direction of the apparent motion. The image size is of 480×480.

As depicted in Fig. 10, the diagonal configuration is not really interesting for small distances between the two templates ($d < 150$ pixels). But, for our fixed distance d of 240 pixels, the use of this configuration is of interest to plainly exploit the available space. Practically, the next displacement $\delta\widetilde{\mathbf{p}}_k$ can be estimated using a Kalman filter or a linear prediction filter[35] minimizing the prediction error in the least squares sense over a set of l previous values of $\delta\mathbf{p}_i$, with $i = k-l, \cdots, k-1$. Hence, to increase the available free space in the motion direction, the configuration parameter α_k can be set according to an estimated value of β_k, denoted $\widetilde{\beta}_k$ and calculated as

$$\widetilde{\beta}_k = \angle\left(\delta\widetilde{\mathbf{p}}_k\right). \tag{19}$$

The commutation angle $\Delta\beta$ between the different orientations (see Fig. 9) is given by

$$\Delta\beta = \arctan\left(d_3/d_2\right). \tag{20}$$

In pseudo-code, α_k can be calculated as described in Algorithm 1, using the function $\text{Round}(x, y)$, the rounding value of x to the nearest multiple of y.

Now, the last configuration parameter r_k can be determined by taking into consideration the vehicle motion, pushing the search area in the direction opposite to

Algorithm 1 α_k configuration.

if ($\left|\widetilde{\beta}_k\right| < \pi/2 - \Delta\beta$ and $\left|\widetilde{\beta}_k\right| > \Delta\beta$) or ($\left|\widetilde{\beta}_k\right| < \pi/2 - \Delta\beta$ and $\left|\widetilde{\beta}_k\right| > \Delta\beta$) **then**
  $\alpha_k = \text{Round}(\widetilde{\beta}_k, \pi/4)$
else
  $\alpha_k = \text{Round}(\widetilde{\beta}_k, \pi/2)$
end if

the vehicle motion and/or the template in the direction of the vehicle motion, creating more available space for template shift and higher velocity accordingly. This method is efficiently exploited in [27] for one template cross-correlation on car-like vehicles. For the proposed two-template correlation, the parameter r_k of the shifted two-template configuration can be calculated as

$$r_k = \|\delta\widetilde{\mathbf{p}}_k\| . \tag{21}$$

At the initialization, the vehicle is at rest and the forthcoming motion direction is obviously unpredictable. Then, the primary template configuration is arbitrarily chosen for the orientation, but the two templates are symmetrically placed about the image center point ($r_k = 0$). To avoid unnecessary switching, the dynamic configuration selection is activated on the only condition that the estimated motion $\delta\widetilde{\mathbf{p}}_k$ exceed a predefined value, which can be set to a percentage of the search area. Under this condition, the two parameters r_k and α_k, defining the two-template configuration, are calculated using (21) and Algorithm 1.

4.2. Pre-rotation of image to increase rotational speed

As reported in section 3.2, the principle of the proposed odometry supposes that the maximum rotation for consecutive images is limited to almost ±4 degrees. However, based on the previous values of the rotation and exploiting the same linear prediction approach as previously mentioned, the estimation of the next rotation of the camera frame $\delta\widetilde{\theta}_k$ can be calculated. The image I_{k-1} may be rotated in advance to limit the rotation with the current frame I_k. The pre-rotated image, noted $\widetilde{I}_{k-1}$, is given by

$$\widetilde{I}_{k-1} = \text{Rotate}\left(I_{k-1}, \delta\widetilde{\theta}_k\right), \tag{22}$$

with $\text{Rotate}(I, \theta)$ the rotated image by angle θ in a counterclockwise direction around the center point of image I. Hence, the maximum angular velocity becomes theoretically unlimited. Now, the limitation is on the maximum angular acceleration value, but it is less restrictive. For instance, the angular acceleration value could reach 3600 degree/s^2 (62.8 rad/s^2) for F_s of 30 Hz.

Finally, the description of the whole visual odometry algorithm, taking account of the proposed adaptations to higher-speed systems, is given in Fig. 11.

5. Experimental results

5.1. Visual odometry system

The ground-based visual odometry approach proposed in this paper has been successfully tested on an industrial omnidirectional mobile robot, the recent KUKA KMR iiwa depicted in Fig 12. The robot has been equipped with a monochromatic camera from Basler company (Ace series acA640-90um USB3 with Sony ICX424 CCD sensor and sensor size of 1/3") associated to a camera lens from Fujinon (9 mm, 2/3"). The maximum frame rate of this compact camera is 90 Hz for an image size of 659 × 494 pixels and the

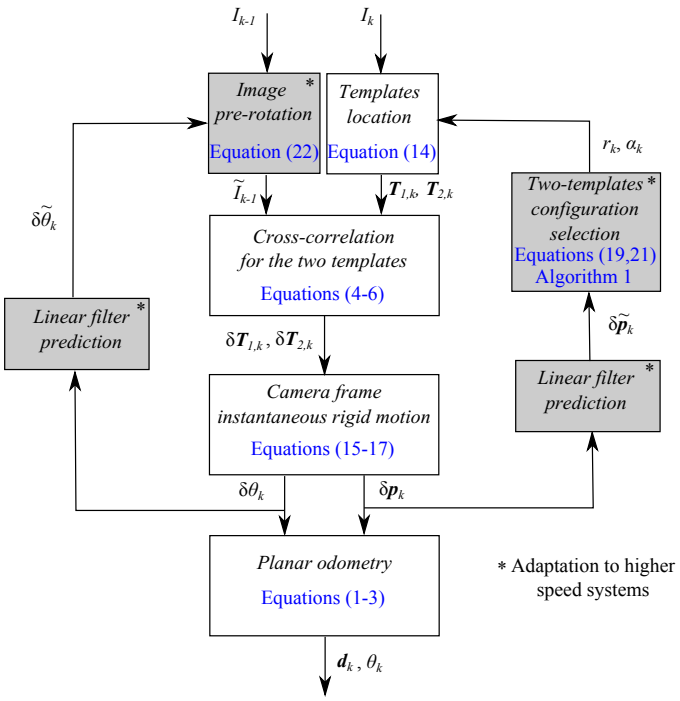

Fig. 11. Flowchart of the proposed visual odometry algorithm.

Fig. 12. Omnidirectional test setup : 4 Mecanum wheels industrial robot (KUKA KMR). The maximum velocity of this autonomous system is 1.4 m/s. For experiments, the robot was remotely operated. The ground truth was measured by a laser tracker and the laser reflector was placed on the top of the CDD camera.

pixel size is 7.4 μm. The acquisition frequency of the camera was set to $F_s = 30$ Hz. The ground truth was measured by a laser tracker, from API Inc., with an absolute accuracy of almost 700 μm /m and a sampling frequency of 250 Hz. The laser tracker was set up at floor level and the laser reflector was placed on the top of the downward loocking camera, which was mounted approximatively 600 mm above ground level. The working area of the mobile robot was limited by the maximum measurement distance of the laser tracker, which is approximately of 16 m. For the experiments, the ground surface was a warehouse concrete floor. The odometry calculation was running on a 64-bits laptop with quad-core Intel Xeon $E5-1620$ CPU (3.60 GHz) and 8GB of RAM and the program was written in C++ using the OpenCV (Open Computer Vision) library (version 2.4.9).

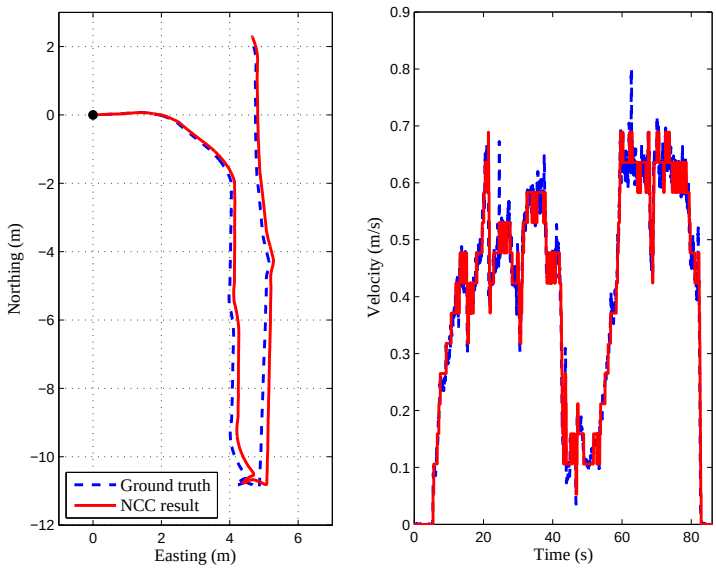

Fig. 13. Odometry result along a 28 m path (left) and associated velocity plot (right).

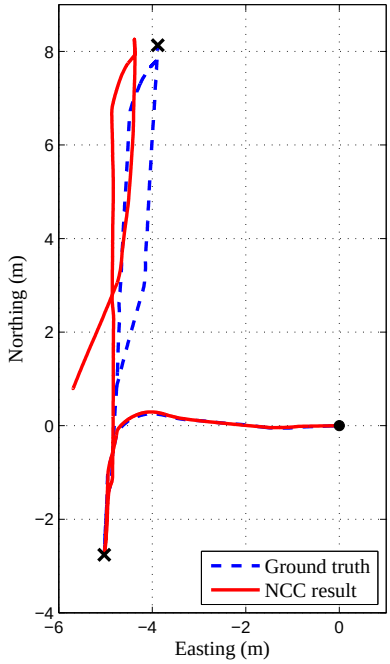

Fig. 14. Odometry result along a 26 m path with two full stop points (cross marks representing for example two workstations). The robot is rotated by 180 degrees about the stop points.

5.2. Accuracy analysis

To present the performance of the proposed odometry module, we use the same approach that is described in [28, 36]. In our case, the first step consists in aligning the beginning

Table II . Visual odometry results for the tests of Fig. 13 and Fig. 14.

Test	Distance error in 10 m
	Median (Std dev σ)
NCC odometry (Fig. 13)	0.11 m ($\sigma = 0.02$)
NCC odometry (Fig. 14)	0.13 m ($\sigma = 0.06$)

part of the odometry path (the first meter) with the ground truth path. Then, the odometry error is calculated as the Euclidean distance between the two path 10 m further. One notes that the 10 m length are measured on the ground truth path and the corresponding point on the odometry path is found using the trajectory timestamp. Lastly, the start position is shifted by 1 m and the previous measure is repeated. From the set of calculated errors, we extract the median and standard deviation values as the performance criterion.
For the experimentation, the robot was driven at a maximum speed of 1 m/s in our machining workshop. Fig. 13 and Fig. 14 present two paths compared to ground truth. Table II shows the accuracy performance calculated as previously defined. We can observe that the overall distance error is below 2 % of the traveling path.

5.3. Some remarks on Optical Flow alternative
In this section, the NCC based method is compared with the Optical-Flow alternative. One of the more popular methods for OF computation in mobile robotics is Lucas and Kanade's local differential technique.[37] In order to allow the tracker to handle larger motion, people usually use a pyramidal implementation of the Lucas Kanade Feature Tracker.[38] Fig. 15 shows a path obtained using the pyramidal Lucas-Kanade method (with a pyramidal level of 5). The same dataset of ground image collected during the test described in Fig. 13 was processed offline. The ground truth path was followed closely during the first 10 meters, but the estimation accuracy was deteriorated in the area enclosed by the dotted line. In this area, the robot velocity was not higher (see the velocity plot in Fig. 13), but the illumination variations caused by a large bay window can explain the local unsuccessful result of the OF method in this area. The correlation-based approach does not suffer from this lack of robustness. Moreover, for the downward facing configuration, the OF method can be very limited regarding the motion amplitude between each image. In,[28] authors report a maximum displacement of almost 15 pixels on different type of surface. Fig. 16 presents an example of estimated displacement for both methods during a linear acceleration stage. For this test, we have used a higher speed vehicle (a car) on asphalt with the camera mounted on the rear bumper. No ground truth information was available for this experiment, but we notice clearly that the OF method can only track displacement up to 60 pixels. The NCC odometer has no such limitation.

6. Conclusion
In this paper, a real-time visual odometry algorithm based on ground appearance and using only one monocular camera is proposed. The planar rigid motion of the vehicle, i.e. position and orientation, is estimated independently of the vehicle steering type, which is perfectly suitable for omnidirectional vehicles. The nearly invariant property of the cross-correlation metric for small rotation and limited template size is highlighted. Hence, a multi-template approach is developed to recover the camera frame orientation. Finally, we proposed a dynamic selection of a two-template configuration to increase the reachable velocity. Experimental tests performed using an omnidirectional industrial platform on indoor surfaces show that the visual odometry method achieved average error rates below 2 % compared to the Ground Truth and gives a coherent framework for practical odometry applications. This visual odometry solution can be used as an efficient and generic substitute or as a complementary system for wheel odometry, or can be integrated in multi-sensor localization frameworks.

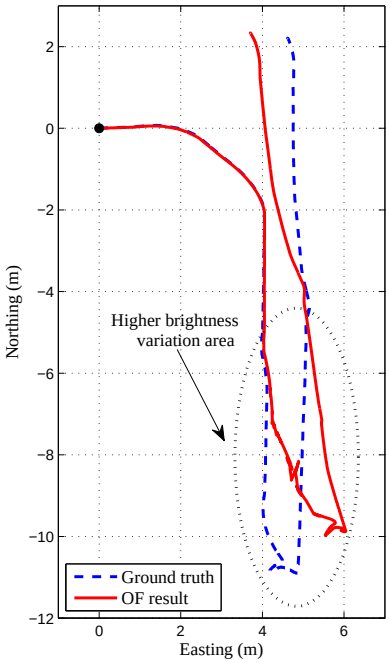

Fig. 15. Odometry result using off-line pyramidal Lucas-Kanade optical flow method on the same dataset of images used in Fig. 13.

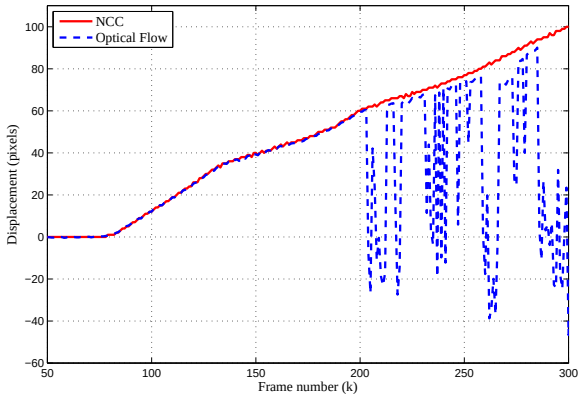

Fig. 16. Displacement estimation (in pixel) of a ground vehicle (passenger car) on asphalt using NCC and OF. The maximum displacement that optical flow can track with our setup is limited to almost 60 pixels (< 8km/h).

References

1. J.A. Batlle, A. Barjau, "Holonomy in mobile robots," *Robotics and Autonomous Systems* **57**(4), 433–440, 2009.
2. K. Jungmin, W. Seungbeom, K. Jaeyong, D. Joocheol, K. Sungshin, B. Sunil, "Inertial navigation system for an automatic guided vehicle with Mecanum wheels," *Int. J. Precis. Eng. Manuf.* **13**(3), 379–386, (2012).

3. A. Garulli, A. Giannitrapani, A. Rossi and A. Vicino, "Mobile robot SLAM for line-based environment representation," **In**: *Proc. IEEE Conference on Decision and Control* (2005), 2041–2046.
4. Z. Riaz, A. Pervez, M. Ahmer and J. Iqbal, "A fully autonomous indoor mobile robot using SLAM," **In**: *Proc. of IEEE Int. Conf. on Inform. and Emerging Technologies*, (2010), 14–16.
5. D. Scaramuzza, F. Fraundorfer, "Visual Odometry: Part I - The First 30 Years and Fundamentals," *IEEE Robotics and Automation Magazine* **18**(4), (2011).
6. A. Comport, E. Malis, P. Rives, "Accurate quadrifocal tracking for robust 3d visual odometry," **In**: *Proc. IEEE Int. Conf. Robotics and Automation*, (2007), 40–45.
7. K. Bernd, A. Geiger and H Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," **In**: *Proc. IEEE Intell. Veh. Symp.*, (2010), 1095–1098.
8. J. Zhang, S. Singh and G. Kantor, "Robust Monocular Visual Odometry for a Ground Vehicle in Undulating Terrain," **In**: *Proc. 8th Int. Conf. on Field and Service Robots*, (2012).
9. D. Scaramuzza, F. Fraundorfer, M. Pollefeys, "Closing the loop in appearance-guided omnidirectional visual odometry by using vocabulary trees," *Robotics and Autonomous Systems* **58**, 820–827,(2010).
10. S. Lovegrove, A.J. Davison and J. Ibanez-Guzman, "Accurate Visual Odometry from a Rear Parking Camera," **In**: *Proc. IEEE Intell. Veh. Symp.*, (2011), pp. 788-793.
11. S. Lee, "Mobile robot localization using optical mice," **In**: *Proc. IEEE Conf. Robot., Automation and Mechatronics*, (2004).
12. S. Kim,S. Lee, "Robustness Analysis of Mobile Robot Velocity Estimation Using a Regular Polygonal Array of Optical Mice," **In**: *Proc. IFAC 17th World Congress*, (2008).
13. A. Bonarini, M. Matteucci and M. Restelli, "A kinematic-independent dead-reckoning sensor for indoor mobile robotics," **In**: *Proc. IEEE /RSJ Int. Conf. on Intell. Rob. Syst. (IROS)*, **4**, 3750–3755, (2004).
14. S. Kim, "Isotropic Optical Mouse Placement for Mobile Robot Velocity Estimation," *Int. J. Adv. Robot. Syst.*, 11–88, (2014).
15. M. Dille,B.P. Grocholsky and S. Singh, "Outdoor Downward-facing Optical Flow Odometry with Commodity Sensors," **In**: *Proc. Int. Conf. on Field and Service Robots*, (2009).
16. M. Irani and P. Anandan, "About Direct Methods, Vision Algorithms: Theory and Practice," *Lecture Notes in Computer Science*, **1883**, 267–277, (2000).
17. D.K. Karna, S. Agarwal and S. Nikam, "Normalized cross-correlation based fingerprint matching," **In**: *Proc. Comput. Graphics, Imaging Visualisation (CGIV)*, (2008), 229–232.
18. Y. Yu, C. Pradalier and G. Zhong, "Appearance-based monocular visual odometry for ground vehicles," **In**: *Proc. IEEE Int. Conf. Adv. Int. Mech. (AIM)*, (2011), 862–867.
19. B.S. Reddy and B.N. Chatterji, "An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration," *IEEE Trans. Image Process.*, **5**(8), 1266–1271, (1996).
20. F. de Morsier, M. Borgeaud, C. Kchler, A. Vogel and V. Gass, "Robust Phase-Correlation Based Registration of Airborne Videos Using Motion Estimation," *Earth Observation of Global Changes (EOGC), Springer Berlin Heidelberg*, (2013).
21. T. Kazik and A.H. Goktogan, "Visual odometry based on the Fourier-Mellin transform for a rover using a monocular ground-facing camera," **In**: *Proc. IEEE Int. Conf. Mech. (ICM)*, (2011), 469–474.
22. J.R. Martinez-de Dios and A. Ollero, "A Real-Time Image Stabilization System Based on Fourier-Mellin Transform," *Lecture Notes in Comput. Science*, **3211**, 376–383, (2002).
23. K. Nagatani, S. Tachibana, M. Sofue and Y. Tanaka, "Improvement of odometry for omnidirectional vehicle using optical flow information," **In**: *Proc. IEEE/RSJ Int. Conf. Intell. Rob. Sys. (IROS)*, (2000), 468–473.
24. A.J. Swank, "Localization Using Visual Odometry and a Single Downward-Pointing Camera," *NASA/TM-2012-216043*, (2012).
25. J. Campbell, R. Sukthankar and I. Nourbakhsh, "Techniques for evaluating optical flow for visual odometry in extreme terrain," **In**: *Proc. IEEE/RSJ Int. Conf. Intell. Rob. Sys. (IROS)*, (2004), 3704–3711.
26. S. Baker and I. Matthews, "Lucas-Kanade20 Years On: A Unifying Framework," *Int. J. of Computer Vision*, **56**(3), 221–255, (2004).
27. N. Nourani-Vatani, J. Roberts and M. V. Srinivasan, "Practical visual odometry for car-like vehicles," **In**: *Proc. IEEE Int. Conf. Rob. Aut. (ICRA*, (2009), 3551–3557.
28. N. Nourani-Vatani and P. V. K. Borges, "Correlation-based visual odometry for ground vehicles," *J. Field Robotics*, **28**(5), 742–768, (2011).
29. R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, **15**(10)5, 1042–1052, (1993).
30. J.N. Sarvaiya, S. Patnaik and S. Bombaywala, "Image Registration by Template Matching Using Normalized Cross-Correlation," **In**: *Proc. Int. Conf. Advances in Computing, Control, Telecommun. Technologies (ACT)*, (2009), 819–822.
31. F. Zhao, Q. Huang and W. Gao, "Image matching by normalized cross-corelation," **In**: *Proc. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, (2006).
32. L. Jianwen and E.E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, **57**(6), 1347–1357, (2010).
33. J. Martin and J.L. Crowley, "Comparison of Correlation Techniques," **In**: *Proc. Conf. Intell. Autonomous Syst. (IAS)*, (1995).

34. E. Arias-Castro and D.L. Donoho, "Does median filtering truly preserve edges better than linear filtering?," *Annals of Statistics*, **37**(3), 1172–2009, (2009).
35. P.P. Vaidyanathan, "The Theory of Linear Prediction," *Morgan and Claypool Publishers*, (2008).
36. S. Johnson, A.E. Goldberg, S.B. Cheng and L.H. Matthies, "Robust and efficient stereo feature tracking for visual odometry," **In**: *Proc. IEEE Int. Conf. Robot. Automat.*, (2008), 11–88.
37. B.D. Lucas and T. Kanade, "An iterative image registration technique with an application tostereo vision," **In**: *Proc. Int. Joint Conf. Artificial Intell.*, (1981), 674–679.
38. J.Y. Bouquet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description," *Tech. Rep. Intel Corporation Microsoft Research Lab*, CA, USA, (2000).