Cost-effective visual odometry system for vehicle motion control in agricultural environments

Published version:

DOI:10.1016/j.compag.2019.03.037

S. Zaman, L. Comba, A. Biglia, D. Ricauda Aimonino, P. Barge, P. Gay

Volume 162, Elsevier, 2019, pag. 82 - 94

Cost-effective visual odometry system for vehicle motion control in agricultural environments

Abstract

In precision agriculture, innovative cost-effective technologies and new improved solutions, aimed at making operations and processes more reliable, robust and economically viable, are still needed. In this context, robotics and automation play a crucial role, with particular reference to unmanned vehicles for crop monitoring and site-specific operations. However, unstructured and irregular working environments, such as agricultural scenarios, require specific solutions regarding positioning and motion control of autonomous vehicles.

In this paper, a reliable and cost-effective monocular visual odometry system, properly calibrated for the localisation and navigation of tracked vehicles on agricultural terrains, is presented. The main contribution of this work is the design and implementation of an enhanced image processing algorithm, based on the cross-correlation approach. It was specifically developed to use a simplified hardware and a low complexity mechanical system, without compromising performance. By providing sub-pixel results, the presented algorithm allows to exploit low resolution images, thus obtaining high accuracy in motion estimation with short computing time. The results, in terms of odometry accuracy and processing time, achieved during the in-field experimentation campaign on several terrains, proved the effectiveness of the proposed method and its fitness for automatic control solutions in precision agriculture applications.

Keywords: Precision agriculture; Visual odometry; Unmanned ground vehicle (UGV); Real- time image processing; Agricultural field robots

Nomenclature

$\mathrm{CEP}_{\varepsilon_s}$	Circular error probable of translation assessment errors [mm]
$d_{\mathrm{i,j}}$	Digital number of pixel located at $\mathrm{i^{th}}$ row and $\mathrm{j^{th}}$ column of image I
$\bar{d}_{\mathrm{u,v}}$	Average values of digital numbers within a portion of image I
$[f_{\mathrm{x}}, f_{\mathrm{y}}]$	x and y component of image focal length [pixel]
g_{x}	Image pixels spatial resolution [mm/pixel]
g_{y}	Image pixels spatial resolution [mm/pixel]
h_{c}	Camera height from the ground [mm]
I_{k}	Acquired grey scale image at time instant t_{k}
$\ell_{\mathrm{i,j}}$	Digital number of pixel located at $\mathrm{i^{th}}$ row and $\mathrm{j^{th}}$ column of image L
$\bar{\ell}$	Average values of digital numbers within template $T(\vartheta)$
$L_{\mathrm{k}}(\vartheta)$	Image obtained by rotating image I_{k} by angle ϑ
n_{Γ}	Distance threshold from γ_{M}
m	Coefficient to set the threshold values for γ
N_{i} x N_{j}	Image size (height x width) [pixel]
$O_{\mathrm{k}}^{\{UGV\}_{\mathrm{k}}}$	Origin of the $\{UGV\}_{\mathrm{k}}$ reference frame at time t_{k}
p_{T}	Template size
$p_{\mathrm{i,j}}^{\{UGV\}_{\mathrm{k}}}$	Position of pixel $d_{\mathrm{i,j}}$ in the reference frame $\{UGV\}_{\mathrm{k}}$ at time t_{k} [mm]
$p_{\hat{u},\hat{v}}^{\{UGV\}_{\mathrm{k+1}}}$	Position of the template $T_{\mathrm{k}}(\hat{\vartheta})$ centre in image $I_{\mathrm{k+1}}$ [mm]
$[p_{\mathrm{c,x}}, p_{\mathrm{c,y}}]^T$	Position coordinates of the camera centre in the $\{UGV\}_{\mathrm{k}}$ reference frame [mm]
$q(u, v, \vartheta)$	Binary function to select a neighbourhood Γ of $\gamma(u, v, \vartheta)$
$R(\cdot)$	Rotation matrix
$s(\cdot)$ (or $s_{\mathrm{k}}^{\mathrm{k+1}}(\cdot)$)	Evaluated vehicle translation (between time instant t_{k} and $t_{\mathrm{k+1}}$) [mm]
s_{r}	Reference vehicle translation [mm]
t_{k}	Generic image acquisition time instant [s]
$T_{\mathrm{k}}(\vartheta)$	Pixel subset, called template, of image $L_{\mathrm{k}}(\vartheta)$
U	Ordered set of u indices
$[\hat{u}_{\mathrm{e}}, \hat{v}_{\mathrm{e}}, \hat{\vartheta}_{\mathrm{e}}]$	Weighted centroid of Γ
$\{UGV\}_{\mathrm{k}}$	Reference frame of the UGV at time t_{k}
V	Ordered set of v indices
w_{T}	Semi-width of the template T_{k} [pixels]

Greek letters

$\gamma(u, v, \vartheta)$	Normalised cross-correlation function
γ_{M}	Maximum value of $\gamma(u, v, \vartheta)$
δ_{ϑ}	Angular resolution of the VO process [deg]
Γ	Specific subset of γ

ε_{s}	Error in translation assessment between two successive images [mm]
ε_{ϑ}	Error in orientation assessment between two successive images [deg]
ϑ	Rotation angle of image $L_{\mathrm{k}}(\vartheta)$ [deg]
$\hat{\vartheta}$	Evaluated vehicle rotation [deg]
ϑ_{r}	Reference vehicle rotation [deg]
ϑ_{min}	Minimum value of $\vartheta \in \Theta$ [deg]
ϑ_{max}	Maximum value of $\vartheta \in \Theta$ [deg]
Θ	Ordered set of all considered rotation angles ϑ ($\Theta = \{\vartheta_{\mathrm{min}}, \vartheta_{\mathrm{min}} + \delta_{\vartheta}, \dots, \vartheta_{\mathrm{max}}\}$) [deg]
$\mu_{\varepsilon_{\vartheta}}$	Average of rotation assessment errors [deg]
$\sigma_{\varepsilon_{\mathrm{s}}}$	Standard deviation of translation assessment errors [mm]
$\sigma_{\varepsilon_{\vartheta}}$	Standard deviation of rotation assessment errors [deg]

Acronyms

CCD	Charged coupled device
CEP	Circular error probable
GPS	Global positioning system
GSD	Ground sample distance
IMU	Inertial measurement unit
NCC	Normalised cross correlation
PA	Precision agriculture
SSD	Sum of squared differences
UGV	Unmanned ground vehicle
VO	Visual odometry

1. Introduction

Precision agriculture (PA) has been recognised as an essential approach to optimise crop-managing practices and to improve field products quality ensuring, at the same time, environmental safety (Ding et al., 2018; Grella et al., 2017; Lindblom et al., 2017). In very large fields and/or in-fields located on hilly areas, cropland monitoring and maintenance may result in a laborious task, requiring automatic machines and procedures (Comba et al., 2018; Grimstad et al., 2017). In this regard, unmanned ground vehicles (UGVs) are playing a crucial role in increasing efficiency in cultivation, e.g. in optimising the use of fertilisers or precision weed control (Utstumo et al., 2018; Vakilian and Massah, 2017; De Baerdemaeker, 2013).

To perform agricultural in-field tasks with the least amount of human interaction, UGVs should be characterised by a high level of automation (van Henten et al., 2013; Kassler, 2001). Nowadays, developed autonomous navigation systems, which use GPS technologies (Bonadies and Gadsden, 2018) and/or machine vision approaches (García-Santillán et al., 2017), allow UGVs, for example, to follow crop rows autonomously, even in complex agricultural scenarios. A common requirement for these applications is a robust up-to-date position and orientation assessment during movements (Ghaleb et al., 2017). Despite the wide diffusion of GPS systems, they show limitations and drawbacks when high precision navigation is required or where the satellite signal is poor, e.g. in covered areas, greenhouses or peculiar hilly regions (Ericson and Åstrand, 2018; Aboelmagd et al., 2013). In agricultural environments, UGV motion estimation by wheel odometry also encounters critical limitations due to wheels slippage on sloped terrains, which is very typical in some crops such as vineyards (Bechar and Vigneault, 2016; Aboelmagd et al., 2013; Nourani-Vatani et al., 2009).

Visual odometry (VO), the measurement of the position and orientation of a system by exploiting the information provided by a set of successive images (Moravec, 1980), can provide reliable movement feedback in UGV motion control (Aqel et al., 2016; Scaramuzza and Fraundorfer, 2011). The hardware required to implement a VO system consists of one or more digital cameras, an image processing unit and an optional lighting system. Not requiring external signals or references, visual odometry has been proven to be very significant in particular contexts where the GPS signal is weak or absent (even where the magnetic field cannot be exploited by compass), by overcoming the limitations of other methodologies (Scaramuzza and Fraundorfer, 2011).

Two main typologies of VO systems can be defined on the basis of the adopted number of cameras: (1) stereo systems use data provided by multiple cameras while (2) monocular systems, characterised by a simple and cost-effective setup, exploit a single digital camera. The image processing of stereo systems is typically complex and time consuming and requires accurate calibration procedures; indeed, an unsynchronised shutter speed between the stereo cameras can lead to errors in motion estimation (Aqel et al., 2016; Jiang et al., 2014). However, the stereo system degrades to the monocular case when the stereo baseline (the distance between the two cameras) is small compared to the distance of the acquired scene by the cameras (Aqel et al., 2016).

The available image processing algorithms for VO applications have two main approaches: (1) feature-based algorithms and (2) appearance-based algorithms. In feature-based VO, specific features/details detected and tracked in the sequence of successive images are exploited (Fraundorfer and Scaramuzza, 2012). Depending on the application, the performance to be achieved and the different approaches in feature selection, several algorithms can be found in literature, such as Libviso (Geiger et al., 2012), Gantry (Jiang et al., 2014) or the Newton-Raphson search methods (Shi and Tomasi, 1994). A different approach is adopted in appearance based-algorithms where successive image frames are searched for changes in appearance by extracting information regarding pixels displacement. The template matching process, which is a widely recognised approach among VO appearance-based solutions, consists in selecting a small portion within a frame (called template) and in comparing it with a temporally subsequent image, then scoring the quality of the matching (Gonzalez et al., 2012; Goshtasby et al., 1984). This task has mainly been performed by using the sum of squared differences (SSD) and normalised cross-correlation (NCC) as similarity measures (Aqel et al., 2016; Yoo et al., 2014; Nourani-Vatani et al., 2009). This latter matching measure, even if computationally heavier than SSD, is invariant to the linear gradient of image contrast and brightness (Mahmood and Khan, 2012; Lewis, 1995).

Motion assessment by VO systems has been proven to be particularly effective when integrated with other sensors such as the inertial measurement unit (IMU), compass sensor, visual compass (Gonzalez et al., 2012), GPS technology or encoders (e.g. on wheels and tracks), to avoid error accumulation on long missions (Zaidner and Shapiro, 2016). Indeed, with particular attention to agricultural applications, innovative and reliable solutions should be developed to reduce system complexity and costs by implementing smart algorithms and by exploiting data fusion (Comba et al., 2016; Zaidner and Shapiro, 2016).

In this paper, a reliable and cost-effective monocular visual odometry system, properly calibrated for the localisation and navigation of tracked vehicles on agricultural terrains, is presented. The main contribution of this work is the design and implementation of an enhanced image processing algorithm, based on the cross-correlation approach, with sub-pixel capabilities. It was specifically developed to use a simplified hardware and a low complexity mechanical system, without compromising performance. In the implemented VO system, installed on a full electric tracked UGV, ground images acquisition was performed by an off-the-shelf camera. The performance of the system, in terms of computing time and of movement evaluation accuracy, was investigated with in-field tests on several kinds of terrains, typical of agricultural scenarios. In addition, the optimal set of algorithm parameters was investigated for the specific UGV navigation/motion control for precision agricultural applications.

The paper is structured as follows: Section 2 reports the description of the implemented tracked UGV and of the vision system. The proposed algorithm for visual odometry is presented in Section 3, while the results from the in-field tests are discussed in Section 4. Section 5 reports the conclusion and future developments.

2. System setup

The implemented VO system was developed to perform the motion and positioning controls of a full electric UGV specifically designed for precision spraying in tunnel crop management, where GPS technology is hampered by metal enclosures. Image acquisition is performed by a Logitech C922 webcam, properly positioned in the front part of the vehicle, with a downward looking setup at the height (h_c) of 245 mm from the ground. To improve the quality of the acquired images, the camera was shielded with a properly sized rigid cover to protect the portion of ground within the camera field of view from direct lighting, thus avoiding irregular lighting and the presence of marked shadows. The illumination of the observed ground surface is provided by a lighting system made of 48 SMD LED 5050 modules (surface-mount device light-emitting diode) with an overall lighting power of more than 1,000 lumens and a power consumption of 8.6 W. Fig. 1 reports the diagram of the VO system setup together with an image of the implemented UGV system.

The image acquisition campaign was conducted on five different terrains (soil, grass, concrete, asphalt and gravel), typical of agricultural environments, in order to assess and quantify the performance of the proposed algorithm. Two datasets of more than 16,000 pairs of grey scale images (8-bit colour representation), at two image resolutions, were processed. Images with a high-resolution have a size of 1280x720 pixels (width and height) while low-resolution ones, which were obtained by down sampling the high resolution ones, are 320x240 pixels (width and height). The sample images at high and low resolution, acquired on five different terrains, are shown in Fig. 2.

A grey scale image I_{k}, acquired at time instant t_{k}, can be defined as an ordered set of digital numbers $d_{\mathrm{i,j}}$ as

$$I_{\mathrm{k}} = \left\{ d_{\mathrm{i,j}} \in [0,1,\dots,255] \vee 1 \leq \mathrm{i} \leq N_{\mathrm{i}}, 1 \leq \mathrm{j} \leq N_{\mathrm{j}} \right\} \quad (1)$$

where i and j are the row and column indices while N_{i} and N_{j} are the numbers of pixels per row and column, respectively.

The intrinsic camera parameters and acquisition settings were evaluated by performing a calibration procedure (Matlab© calibration toolbox). The focal length in pixel was $(f_{\mathrm{x}}, f_{\mathrm{y}}) = (299.4122, 299.4303)$ and $(f_{\mathrm{x}}, f_{\mathrm{y}}) = (888.5340, 888.8749)$ for the low-resolution and high-resolution images respectively. The position [mm] of pixels $d_{\mathrm{i,j}}$ in the UGV reference frame $\{UGV\}_{\mathrm{k}}$ at time t_{k}, defined with origin O_{k} in the barycentre of the tracked system and with the x-axis aligned to the vehicle's forward motion direction (Fig. 4), can thus be easily computed as

$$p_{\mathrm{i,j}}^{\{UGV\}_{\mathrm{k}}} = \left[\left(\mathrm{j} - \left\lceil\frac{N_{\mathrm{j}}}{2}\right\rceil\right)\frac{h_{\mathrm{c}}}{f_{\mathrm{x}}}, \left(\left\lceil\frac{N_{\mathrm{i}}}{2}\right\rceil - \mathrm{i}\right)\frac{h_{\mathrm{c}}}{f_{\mathrm{y}}}\right]^{\mathrm{T}} + \left[p_{\mathrm{c,x}}, p_{\mathrm{c,y}}\right]^{\mathrm{T}} \tag{2}$$

where $\frac{h_{\mathrm{c}}}{f_{\mathrm{x}}}$ and $\frac{h_{\mathrm{c}}}{f_{\mathrm{y}}}$ are the pixels' spatial resolutions g_{x} and g_{y} [mm/pixel] respectively and $\left[p_{\mathrm{c,x}}, p_{\mathrm{c,y}}\right]^{\mathrm{T}}$ are the position coordinates of the camera centre [mm] in the $\{UGV\}_{\mathrm{k}}$. In the implemented UGV, the position coordinates of the camera with respect to the barycentre of the tracked system are $[950,0]^{\mathrm{T}}$ mm. The relevant camera and images intrinsic parameters adopted in this work are summarised in Table 1.

3. Visual odometry algorithms

In visual odometry, the objective of measuring the position and orientation of an object at time $t_{\mathrm{k+1}}$, knowing its position and orientation at time t_{k}, is performed by evaluating the relative movement of a solid camera having occurred during time interval $t_{\mathrm{k+1}} - t_{\mathrm{k}}$. This task is performed by comparing the image pair I_{k} and $I_{\mathrm{k+1}}$, acquired in the ordered time instants t_{k} and $t_{\mathrm{k+1}}$, respectively.

In the normalised cross-correlation (NCC) approach, a pixel subset $T_{\mathrm{k}}(\vartheta)$ (also named template) is selected from the image $L_{\mathrm{k}}(\vartheta)$ centre, which is obtained rotating image I_{k} by an angle ϑ, as

$$T_{\mathrm{k}}(\vartheta) = \left\{\ell_{\mathrm{i,j}} \in L_{\mathrm{k}}(\vartheta) \middle| \left|\mathrm{i} - \left\lceil\frac{N_{\mathrm{i}}}{2}\right\rceil\right| \le w_T, \left|\mathrm{j} - \left\lceil\frac{N_{\mathrm{j}}}{2}\right\rceil\right| \le w_T\right\} \tag{3}$$

where $\ell_{\mathrm{i,j}}$ is a digital number of image L_{k} and w_{T} is the semi-width [pixels] of the template T_{k}. The adopted template size p_{T} can be defined as a fraction of the shortest image dimension as $p_{\mathrm{T}} = 2 \cdot w_{\mathrm{T}} \cdot N_{\mathrm{i}}^{-1}$; with this definition $p_{\mathrm{T}} \subset [0\ 1]$. With no assumption on the performed movement, angle ϑ is usually selected from an ordered set of values $\Theta = \{\vartheta_{\min}, \vartheta_{\min} + \delta_{\vartheta}, \dots, \vartheta_{\max}\}$, with

ϑ_{min} and ϑ_{max} chosen to consider the whole circle angle. The δ_ϑ parameter can be defined as the angular resolution of the process.

The relative movement of $I_{\mathrm{k+1}}$ with respect to image I_{k}, in terms of translation $[\hat{u},\hat{v}]^{\mathrm{T}}$ [pixels] and rotation $\hat{\vartheta}$ [deg], is thus performed by assessing the position of the ground portions represented in templates $T_{\mathrm{k}}(\vartheta)$ in the subsequent image $I_{\mathrm{k+1}}$ by solving the problem

$$\gamma_{\mathrm{M}} = \max_{\hat{u},\hat{v},\hat{\vartheta}} \gamma(u,v,\vartheta) \tag{4}$$

with $u \in U = \{w_{\mathrm{T}}, w_{\mathrm{T}}+1, \dots, N_{\mathrm{i}}-w_{\mathrm{T}}\}$, $v \in V = \{w_{\mathrm{T}}, w_{\mathrm{T}}+1, \dots, N_{\mathrm{j}}-w_{\mathrm{T}}\}$, $\vartheta \in \Theta$ and where $\gamma(u,v,\vartheta)$ is the normalised cross-correlation function (Aqel et al., 2016; Lewis, 1995) defined as

$$\gamma(u,v,\vartheta) = \frac{\sum_{\mathrm{i}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \sum_{\mathrm{j}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \left(d_{\mathrm{i}+u,\mathrm{j}+v} - \bar{d}_{u,v}\right)_{I_{\mathrm{k+1}}} \cdot \left(\ell_{\mathrm{i}+w_{\mathrm{T}},\mathrm{j}+w_{\mathrm{T}}} - \bar{\ell}\right)_{T_{\mathrm{k}}(\vartheta)}}{\sqrt{\sum_{\mathrm{i}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \sum_{\mathrm{j}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \left(d_{\mathrm{i}+u,\mathrm{j}+v} - \bar{d}\right)_{I_{\mathrm{k+1}}}^{2} \cdot \left(\ell_{\mathrm{i}+w_{\mathrm{T}},\mathrm{j}+w_{\mathrm{T}}} - \bar{\ell}\right)_{T_{\mathrm{k}}(\vartheta)}^{2}}} \tag{5}$$

with

$$\bar{d}_{u,v} = \frac{\sum_{\mathrm{i}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \sum_{\mathrm{j}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \left(d_{\mathrm{i}+u,\mathrm{j}+v}\right)_{I_{\mathrm{k+1}}}}{4 \cdot w_{\mathrm{T}}^{2}} \tag{6}$$

and

$$\bar{\ell} = \frac{\sum_{\mathrm{i}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \sum_{\mathrm{j}=-w_{\mathrm{T}}}^{w_{\mathrm{T}}} \left(\ell_{\mathrm{i}+w_{T},\mathrm{j}+w_{T}}\right)_{T_{\mathrm{k}}(\vartheta)}}{4 \cdot w_{\mathrm{T}}^{2}} \tag{7}$$

the average values of the digital numbers within a portion of image $I_{\mathrm{k+1}}$ and template $T_{\mathrm{k}}(\vartheta)$, respectively. A scheme of the implemented NCC algorithm is reported in Fig. 3.

The relative movement $s_{\mathrm{k}}^{\mathrm{k+1}}$ performed by the UGV in the time interval $t_{\mathrm{k+1}} - t_{\mathrm{k}}$ (Fig. 4) can thus be easily computed as

$$s_{\mathrm{k}}^{\mathrm{k+1}}(\hat{u},\hat{v},\hat{\vartheta}) = R\left(-\hat{\vartheta}\right) \cdot p_{\hat{u},\hat{v}}^{\{UGV\}_{\mathrm{k+1}}} - p_{\left\lceil\frac{N_{\mathrm{i}}}{2}\right\rceil,\left\lceil\frac{N_{\mathrm{j}}}{2}\right\rceil}^{\{UGV\}_{\mathrm{k}}} \tag{8}$$

where $R\left(-\hat{\vartheta}\right)$ is the rotation matrix of angle $-\hat{\vartheta}$, $p_{\hat{u},\hat{v}}^{\{UGV\}_{\mathrm{k+1}}}$ is the template $T_{\mathrm{k}}\left(\hat{\vartheta}\right)$ assessed position [mm] in $I_{\mathrm{k+1}}$ (represented in $\{UGV\}_{\mathrm{k+1}}$, Eq. (2)), and $p_{\left\lceil\frac{N_{\mathrm{i}}}{2}\right\rceil,\left\lceil\frac{N_{\mathrm{j}}}{2}\right\rceil}^{\{UGV\}_{\mathrm{k}}}$ is the known position [mm] of template T_{k} in I_{k}, (represented in $\{UGV\}_{\mathrm{k}}$, Eq. (2) . For the sake of clarity, it should be noted that $p_{\left\lceil\frac{N_{\mathrm{i}}}{2}\right\rceil,\left\lceil\frac{N_{\mathrm{j}}}{2}\right\rceil}^{\{UGV\}_{\mathrm{k}}}$ is equal to $\left[p_{\mathrm{c,x}}, p_{\mathrm{c,y}}\right]^{\mathrm{T}}$, which is $[950{,}0]^{\mathrm{T}}$ millimetres, and that $s_{\mathrm{k}}^{\mathrm{k+1}}(\hat{u},\hat{v},\hat{\vartheta})$ coincides with $O_{\mathrm{k}+1}^{\{UGV\}_{\mathrm{k}}}$, which is the origin of the reference frame $\{UGV\}_{\mathrm{k}+1}$ represented in $\{UGV\}_{\mathrm{k}}$ reference frame (Fig. 4).

3.1 Enhanced cross-correlation algorithm

The quality of the UGV's movement measure, using normalised cross-correlation-based visual odometry algorithms, is strictly related to the solution of the problem defined in Eq. (4). The approach of considering the sole maximum value γ_{M} of $\gamma(u, v, \vartheta)$, with $u \in \{w_{\mathrm{T}}, w_{\mathrm{T}} + 1, \dots, N_{\mathrm{i}} - w_{\mathrm{T}}\}$, $v \in \{w_{\mathrm{T}}, w_{\mathrm{T}} + 1, \dots, N_{\mathrm{j}} - w_{\mathrm{T}}\}$ and $\vartheta \in \Theta$, has intrinsic limitations regarding maximum achievable accuracy. Indeed, the digital discretisation of the field of view performed by the digital camera and the discrete set Θ of the investigated orientation ϑ affect both the translation and the rotation assessments. The accuracy of the VO system is thus related to the adopted image resolution, being directly related to the pixels ground sample distance (GSD) g_{x} and g_{y} and the angle step δ_{ϑ} adopted in the image processing. Regarding this aspect, an accuracy improvement can be pursued by adopting high-resolution cameras, which can provide images with smaller pixels GSD g_{x} and g_{y}: favourable effects are linked, in the meanwhile, to the accuracy of $[\hat{u}, \hat{v}]^{\mathrm{T}}$ and to the angular resolution δ_{ϑ} values. Indeed, concerning the rotation procedure of image $L_{\mathrm{k}}(\delta_{\vartheta})$, if the rotation angle δ_{ϑ} is small, no modifications are obtained on the pixels' digital number in the central part of the image, where the template is selected. For the sake of clarity, the smallest δ_{ϑ} values which lead to template $T_{\mathrm{k}}(\delta_{\vartheta})$ modifications, in relation to image resolution and template size p_{T}, are reported in Table 2.

However, increasing image resolution leads to a considerable increment in the required computing load, which does not fit with the real-time requirements of the VO algorithm application or requires technologies which are too expensive.

The proposed approach is aimed at increasing VO assessment accuracy by using very low-resolution images, which allows to drastically reduce the computing load while achieving results comparable to the ones obtained by processing high-resolution data. This translates into more cost-effective systems, requiring economical acquisition and processing hardware.

For this purpose, a function $q(u, v, \vartheta)$ was defined as

$$q(u,v,\vartheta) = \begin{cases} 0 & \text{if} \quad \gamma(u,v,\vartheta) < m \cdot \gamma_{\mathrm{M}}, \quad \left\| \left([u,v,\vartheta] - [\hat{u},\hat{v},\hat{\vartheta}] \right) \circ [1,1,\delta_{\vartheta}^{-1}] \right\|_2 > n_{\Gamma} \\ 1 & \text{if} \quad \gamma(u,v,\vartheta) \geq m \cdot \gamma_{\mathrm{M}}, \quad \left\| \left([u,v,\vartheta] - [\hat{u},\hat{v},\hat{\vartheta}] \right) \circ [1,1,\delta_{\vartheta}^{-1}] \right\|_2 \leq n_{\Gamma} \end{cases} \tag{9}$$

in order to consider a neighbourhood Γ of the maximum γ_{M} (Eq. (4)) of cross-correlation discrete function $\gamma(u,v,\vartheta)$ in the space (u,v,ϑ), with values higher than $m \cdot \gamma_{\mathrm{M}}$. In particular, n_{Γ} is the distance threshold from γ_{M} and m is the coefficient to set the γ values threshold. In this work, adopted values are $n_{\Gamma} = 5$ and $m = 0.95$ on the base of empirical evaluations. The Hadamard product with $[1,1,\delta_{\vartheta}^{-1}]$ was adopted to normalise the weight of the three spatial coordinates (u,v,ϑ).

The enhanced movement assessment is thus performed by computing the weighted centroids $[\hat{u}_{\mathrm{e}},\hat{v}_{\mathrm{e}},\hat{\vartheta}_{\mathrm{e}}]$ of Γ (Fig. 5), as

$$\hat{u}_{\mathrm{e}} = \frac{\sum_{u=w_{\mathrm{T}}}^{N_{\mathrm{i}}-w_{\mathrm{T}}} u \cdot \sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} \sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} \gamma(u,v,\vartheta_{\mathrm{z}}) \cdot q(u,v,\vartheta_{\mathrm{z}})}{\sum_{u=w_{\mathrm{T}}}^{N_{\mathrm{i}}-w_{\mathrm{T}}} \sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} \sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} q(u,v,\vartheta_{\mathrm{z}})} \tag{10}$$

$$\hat{v}_{\mathrm{e}} = \frac{\sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} v \cdot \sum_{u=1}^{N_{\mathrm{i}}-w_{\mathrm{T}}} \sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} \gamma(u,v,\vartheta_{\mathrm{z}}) \cdot q(u,v,\vartheta_{\mathrm{z}})}{\sum_{u=w_{\mathrm{T}}}^{N_{\mathrm{i}}-w_{\mathrm{T}}} \sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} \sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} q(u,v,\vartheta_{\mathrm{z}})} \tag{11}$$

and

$$\hat{\vartheta}_{\mathrm{e}} = \frac{\sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} \mathrm{z} \cdot \sum_{u=w_{\mathrm{T}}}^{N_{\mathrm{i}}-w_{\mathrm{T}}} \sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} \gamma(u,v,\vartheta_{\mathrm{z}}) \cdot q(u,v,\vartheta_{\mathrm{z}})}{\sum_{u=w_{\mathrm{T}}}^{N_{\mathrm{i}}-w_{\mathrm{T}}} \sum_{v=w_{\mathrm{T}}}^{N_{\mathrm{j}}-w_{\mathrm{T}}} \sum_{\mathrm{z}=1}^{\mathrm{card}(\Theta)} q(u,v,\vartheta_{\mathrm{z}})} \tag{12}$$

With the proposed approach, the UGV's movement evaluation is not defined by discrete values, since $[\hat{u}_{\mathrm{e}},\hat{v}_{\mathrm{e}},\hat{\vartheta}_{\mathrm{e}}] \in \mathbb{R}^3$.

4. Results and discussion

The performance of the proposed visual odometry system, developed for a UGV motion estimation, was assessed by processing more than 16,000 images. The in-field tests were performed on different agricultural terrains by acquiring images on soil, grass, asphalt, concrete and gravel. In particular, both rectilinear and curvilinear paths were planned. Considering the whole dataset, the travelled distance between two subsequent images ranges between 0 mm (static vehicle) and 70 mm, which guarantees a minimum overlapping area of 72%. The relative rotation does not exceed the range of [-9 +9] degrees, due to the short movement between two acquired frames. The image resolutions were 1280x720 pixels (high-resolution images) and 320x240 pixels (low-resolution images). To evaluate the performance improvements of the proposed algorithm, with sub-pixel capabilities, the set of acquired images was also processed by means of a standard VO algorithm (Computer Vision System Toolbox, MathWorks, 2018).

The performance analysis of the proposed VO system was performed: (1) by assessing motion evaluation accuracy in pairs of successive images, using high-resolution datasets as a reference, and (2) by computing the cumulative error with respect to in-field position references travelling about 10 meters long paths.

Concerning a pair of successive images, the error in measuring the relative movement s and the rotation ϑ between two subsequent images was defined as

$$\varepsilon_{\mathrm{s}} = \|s(\cdot) - s_{\mathrm{r}}\|_2 \qquad (13)$$

and

$$\varepsilon_{\vartheta} = \left|\hat{\vartheta} - \vartheta_{\mathrm{r}}\right| \qquad (14)$$

respectively, where $s(\cdot)$ (Eq. (8)) and $\hat{\vartheta}$ are the vehicle's movement and rotation, evaluated by using the enhanced and standard algorithm and by processing low-resolution images, while s_{r} and ϑ_{r} represent the reference measurements from the high-resolution images. Concerning the translation assessment, accuracy was expressed by the circular error probable ($\mathrm{CEP}_{\varepsilon_{\mathrm{s}}}$) and standard deviation ($\sigma_{\varepsilon_{\mathrm{s}}}$) indices (Winkler et al., 2012) (Table 3), while accuracy in measuring the changes in vehicle orientation ϑ were described by computing the average ($\mu_{\varepsilon_{\vartheta}}$) and standard deviation ($\sigma_{\varepsilon_{\vartheta}}$) of the computed ε_{ϑ} errors (Table 4). The results were detailed for each in-field test performed on a specific kind of terrain and, finally, computed by considering the whole image dataset. Overall accuracy in the translation assessment of the proposed algorithm across different terrains resulted to be $CEP_{\varepsilon_{\mathrm{s}}} = 0.16$ mm, with an improvement of around 54% with respect to the values obtained by processing the images with the standard algorithm, which shows a $CEP_{\varepsilon_{\mathrm{s}}}$ of 0.37 mm. The average error in the vehicle's orientation assessment was $\mu_{\varepsilon_{\vartheta}} = 0.26$ degrees, with an improvement of around 67.6% with respect to the values obtained by processing the images with the standard algorithm. The typology of terrain slightly affects the achieved performance: on the grass surface, a lower performance improvement was found compared to other terrains. Indeed, the greater variability in object height within the camera field of view can lead to additional perspective errors. Nevertheless, even in these complex scenarios, improvements of 44% in the $CEP_{\varepsilon_\mathrm{s}}$ and of 34% in the orientation assessment was observed ($CEP_{\varepsilon_\mathrm{s}}$= 0.19 mm and $\mu_{\varepsilon_\vartheta} = 0.42$ degree) compared to the ones obtained by the standard algorithm. Boxplots of errors ε_s and ε_ϑ, computed by considering the whole image dataset, are reported in Fig. 6 for standard and enhanced algorithms. The x and y components of ε_s and the CEP_{ε_s} circles are detailed in Fig. 7, with ε_ϑ represented by using a colour bar.

The cumulative error was computed for 20 sample paths of the tracked vehicle with a length of 9.6 meters, defined as a curvilinear path generated by a sinusoidal trajectory of 0.15 m amplitude and of 3.2 m period. The number of acquired images for a path repetition ranges between 156 and 166, with an average travelled distance between two consecutive frames of 61 mm. Defining a normalised cumulative error with respect to the travelled distance, the obtained values are 0.08 and 0.84 $[\mathrm{deg} \cdot \mathrm{m}^{-1}]$ for what concerns translation and orientation, respectively. The improvement compared to the standard algorithm is of about 60% for both the translation and orientation assessments. The boxplots of all the obtained cumulative errors, expressed in normalised values, are reported in Fig. 8. Considering a constant travelled distance, the cumulative error is strictly related to the number of processed images, as every processing step contributes to the overall error. With this assumption, to minimise the cumulative error, pairs of frames acquired at the largest distance, still guaranteeing the proper overlapping surface, should be used. For this purpose, a multi-frame approach can further improve system performance (Jiang et al., 2014).

The optimal configuration for a VO system setup requires thorough analysis of the parameters related to image processing and their tuning according to the application requirements. With particular attention to the overall VO system performance, the size p_T of the template $T_\mathrm{k}(\vartheta)$ is a relevant algorithm parameter since it is strictly related to (1) the motion accuracy measure, (2) the allowed maximum length of the relative movement between two subsequent images, which should still assure the required overlapping surface of the template, (3) the computing time and, thus, (4) the maximum allowed velocity with a specific VO setup.

The template size p_T has a non-linear and non-monotonic effect on the overall VO system's accuracy. Considering the translation assessment, by varying p_T within the range 0.05-0.35, an optimal value can be found that provides the best accuracy. Indeed, the proposed algorithm achieves a $CEP_{\varepsilon_\mathrm{s}} = 0.16$ mm for $p_\mathrm{T} = 0.20$, while accuracy degrades to $CEP_{\varepsilon_\mathrm{s}} = 0.21$ mm and $CEP_{\varepsilon_\mathrm{s}} = 0.22$ for $p_\mathrm{T} = 0.05$ and $p_\mathrm{T} = 0.35$, respectively. The boxplots of errors ε_s and ε_ϑ, obtained by setting p_T within the range 0.05-0.35, are reported in Figs. 9 and 10, respectively. The observed accuracy trend in determining the vehicle's orientation is similar to the one described for translation, with the exception of the effect of p_T values greater than 0.20 on the accuracy's decrement: it is less marked until p_T exceeds 0.6, values that lead to insufficient overlapping surfaces between two successive images. Indeed, regarding proper overlapping surfaces between successive images, the template size should not exceed a certain value. Larger template sizes p_T require a shorter relative movement of the vehicle between image acquisition time instants to avoid complete mismatch between a pair of successive images. In the implemented VO system performance evaluation, increasing p_T from 0.1 to 0.6 will limit the maximum allowed movement from 93.1 to 39.2 mm, requiring a higher framerate to keep proper image acquisition when considering a constant vehicle velocity.

Concerning the computing time, smaller p_T values allow to drastically reduce the required time to process an image pair: considering a low-resolution dataset, the average computing time (0.02 seconds) using $p_\mathrm{T} = 0.05$ is 88% shorter than the one required by $p_\mathrm{T} = 0.35$ (0.19 seconds). Fig. 11a reports the average computing time obtained for processing low and high resolution images with a template size p_T ranging from 0.05 to 0.8.

Consequently, the allowed maximum velocity of the vehicle is thus strictly related to template size: considering a constant computing power, smaller template sizes lead to higher vehicle maximum speeds, due to the concurrent effects on the processing time required for an image pair and the length of the maximum allowed movement between two subsequent images. In the implemented VO system, processing low-resolution images by using a value of $p_\mathrm{T} = 0.05$, the upper limit velocity (about $4.1\ \mathrm{m \cdot s^{-1}}$) is 91% greater than the one allowed by $p_\mathrm{T} = 0.35$ (about $0.3\ \mathrm{m \cdot s^{-1}}$). The maximum allowed velocities for low and high-resolution images with respect to template size p_T ranging from 0.05 to 0.8 are represented in Fig. 11b.

5. Conclusions

In this paper, an enhanced image processing algorithm for a cost-effective monocular visual odometry system, aimed at obtaining highly reliable results at low computational costs for a tracked UGV navigation in agricultural applications, is presented. The implemented VO system consists of a downward looking low cost web-camera sheltered with a rigid cover to acquire images with uniform LED lighting. Based on the normalised cross-correlation methodology, the proposed VO algorithm was developed to exploit low-resolution images (320x240 pixels), achieving sub-pixel accuracy in motion estimation. The algorithm allows the VO system to be applied to real-time applications using cost-effective hardware, by requiring a lower computational load.

The robustness of the proposed VO algorithm was evaluated by performing an extensive in-field test campaign on several terrains typical of agricultural scenarios: soil, grass, concrete, asphalt and gravel. The relationship between system performances and more relevant algorithm parameters was investigated in order to determine a proper final system setup.

The obtained overall accuracy, in terms of circular probable error and normalised cumulative error, which are 0.16 mm and 0.08 respectively, were compatible with UGV requirements for precision agricultural applications. The obtained short computing time allowed the vehicle to achieve a maximum velocity limit higher than $4\ \mathrm{m} \cdot \mathrm{s}^{-1}$.

Based on the relative motion assessment, the performance of VO systems degrades when incrementing path length. Therefore, the system integration with absolute reference is required to maintain the needed accuracy during long mission paths.

Acknowledgments

The presented research was partially funded by the Italian Ministry of University and Research PRIN 2015 "Ottimizzazione di macchine operatrici attraverso l'analisi del profilo di missione per un'agricoltura più efficiente" (Prot. 2015KTY5NW) and the "ElectroAgri" Project, Manunet ERA-NET, Call 2016, P.O.R. FESR 2014-2020 European Program.

References

N. Aboelmagd, T.B. Karmat, J Georgy. Fundamentals of inertial navigation, satellite-based positioning and their integration. Springer, (2013), doi:10.1007/978-3-642-30466-8

M.O.A. Aqel, M.H. Marhaban, M.I. Saripan, N.Bt. Ismail. Review of visual odometry: types, approaches, challenges, and applications. SpringerPlus, 5 (2016), doi:10.1186/s40064-016-3573-7

J. De Baerdemaeker. Precision Agriculture Technology and Robotics for Good Agricultural Practices. IFAC Proceedings Volumes, 46 (2013), pp. 1-4, doi:10.3182/20130327-3-JP-3017.00003

A. Bechar, C. Vigneault. Agricultural robots for field operations: Concepts and components. Biosyst Eng, 149 (2016), pp. 94-111, doi:10.1016/j.biosystemseng.2016.06.014

S. Bonadies, S.A. Gadsden. An overview of autonomous crop row navigation strategies for unmanned ground vehicles. Engineering in Agriculture, Environment and Food, 12 (2019), pp. 24-31, doi:10.1016/j.eaef.2018.09.001

L. Comba, A. Biglia, D. Ricauda Aimonino, P. Gay. Unsupervised detection of vineyards by 3D point-cloud UAV photogrammetry for precision agriculture. Comput Electron Agr, 155 (2018), pp. 84-95, doi:10.1016/j.compag.2018.10.005

L. Comba, P. Gay, D. Ricauda Aimonino. Robot ensembles for grafting herbaceous crops. Biosyst Eng, 146 (2016), pp. 227-239, doi:10.1016/j.biosystemseng.2016.02.012

Y. Ding, L. Wang, Y. Li, D. Li. Model predictive control and its application in agriculture: A review. Comput Electron Agr, 151 (2018), pp. 104-117, doi:10.1016/j.compag.2018.06.004

S.K. Ericson, B.S. Åstrand. Analysis of two visual odometry systems for use in an agricultural field environment. Biosyst Eng, 166 (2018), pp. 116-125, doi:10.1016/j.biosystemseng.2017.11.009

F. Fraundorfer, D. Scaramuzza. Visual odometry: Part II: Matching, robustness, optimization, and applications. IEEE Robotics and Automation Magazine, 19 (2012), pp. 78-90, doi:10.1109/MRA.2012.2182810

I.D. García-Santillán, M. Montalvo, J.M. Guerrero, G. Pajares. Automatic detection of curved and straight crop rows from images in maize fields. Biosyst Eng, 156 (2017), pp. 61-79, doi:10.1016/j.biosystemseng.2017.01.013

A. Geiger, P. Lenz, R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), (2012)

F.A. Ghaleb, A. Zainala, M.A. Rassam, A. Abraham. Improved vehicle positioning algorithm using enhanced innovation-based adaptive Kalman filter. Pervasive and Mobile Computing, 40 (2017), pp. 139-155, doi:10.1016/j.pmcj.2017.06.008

R. Gonzalez, F. Rodriguez, J.L. Guzman, C. Pradalier, R. Siegwart. Combined visual odometry and visual compass for off-road mobile robots localization. Robotica, 30 (2012), pp. 865-878, doi:10.1017/S026357471100110X

A. Goshtasby, S.H. Gage, J.F. Bartholic. A two-stage cross correlation approach to template matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (1984), pp. 374-378, doi:10.1109/TPAMI.1984.4767532

M. Grella, E. Gil, P. Balsari, P. Marucco, M. Gallart. Advances in developing a new test method to assess spray drift potential from air blast sprayers. Span J of Agric Res 15 (2017), doi:10.5424/sjar/2017153-10580

L. Grimstad, P.J. From. Thorvald II - a Modular and Re-configurable Agricultural Robot. IFAC-PapersOnLine, 50 (2017), pp. 4588-4593, doi:10.1016/j.ifacol.2017.08.1005

D. Jiang, L. Yang, D. Li, F. Gao, L. Tian, L. Li. Development of a 3D ego-motion estimation system for an autonomous agricultural vehicle. Biosyst Eng, 121 (2014), pp. 150-159, doi:10.1016/j.biosystemseng.2014.02.016

M. Kassler. Agricultural Automation in the new Millennium. Comput Electron Agr, 30 (2001), pp. 237-240, doi:10.1016/S0168-1699(00)00167-8

J.P. Lewis. Fast Template Matching. Vis. Interface, 95 (1995), pp. 120-123

J. Lindblom, C. Lundström, M. Ljung, A. Jonsson. Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. Precis Agric, 18 (2017), pp. 309-331, doi:10.1007/s11119-016-9491-4

A. Mahmood, S. Khan. Correlation-coefficient-based fast template matching through partial elimination. IEEE Transactions on Image Processing, 21 (2012), pp. 2099-2108, doi:10.1109/TIP.2011.2171696

MathWorks (2018). Computer Vision System Toolbox

H. Moravec. Obstacle Avoidance and Navigation in the Real world by a seeing robot rover. PhD thesis, (1980), Stanford University.

N. Nourani-Vatani, J. Roberts, M.V. Srinivasan. Practical Visual Odometry for Car-like Vehicles. IEEE International Conference on Robotics and Automation, 1-7 (2009), pp. 3551-3557, doi:10.1109/ROBOT.2009.5152403

D. Scaramuzza, F. Fraundorfer. Visual Odometry Part I: The First 30 Years and Fundamentals. IEEE Robotics & Automation Magazine, 18 (2011), pp. 80-92, doi:10.1109/MRA.2011.943233

J. Shi, C. Tomasi. Good features to track. EEE Conference on Computer Vision and Pattern Recognition, (1994), doi:10.1109/CVPR.1994.323794

T. Utstumo, F. Urdal, A. Brevik, J. Dørum, J. Netland, Ø. Overskeid et al. Robotic in-row weed control in vegetables. Comput Electron Agr, 154 (2018), pp. 36-45, doi:10.1016/j.compag.2018.08.043

E.J. van Henten, C.W. Bac, J. Hemming, Y. Edan. Robotics in protected cultivation. IFAC Proceedings Volumes, 46 (2013), pp. 170-177, doi:10.3182/20130828-2-SF-3019.00070

K.A. Vakilian, J. Massah. A farmer-assistant robot for nitrogen fertilizing management of greenhouse crops. Comput Electron Agr, 139 (2017), pp. 153-163, doi:10.1016/j.compag.2017.05.012

V. Winkler, B. Bickert. Estimation of the circular error probability for a Doppler-Beam-Sharpening-Radar-Mode. 9th European Conference on Synthetic Aperture Radar, (2012), pp. 368-371

J. Yoo, S.S. Hwang, S.D. Kim, M.S. Ki, J. Cha. Scale-invariant template matching using histogram of dominant gradients. Pattern Recognit, 47 (2014), pp. 3006-3018, doi:0.1016/j.patcog.2014.02.016.

G. Zaidner, A. Shapiro. A novel data fusion algorithm for low-cost localisation and navigation of autonomous vineyard sprayer robots. Biosyst Eng, 146 (2016), pp. 133-148, doi:10.1016/j.biosystemseng.2016.05.002