Dual-Hemispherical Photometric Visual Servoing

Nathan Crombez, Jocelyn Buisson, Antoine N André, Guillaume Caron

Dual-Hemispherical Photometric Visual Servoing

Nathan Crombez[1], Jocelyn Buisson[1], Antoine N. André[2], Guillaume Caron[2,3]

***Abstract*— It is well established that wide field of view cameras, as well as using the whole photometric information contained in images, offer many advantages for visual servoing. Therefore, we propose to extend the photometric visual servoing to full spherical cameras. More precisely, we are dealing with 360-degree optical rigs composed of two wide-angle lenses oriented in opposite directions that capture everything around the device in one acquisition. The photometric visual feature coupled to dual-hemispherical acquisitions that contain the whole surrounding scene provide useful complementary information, showing large convergence domains, straighter camera trajectories than with a single hemispherical camera, and high accuracy. We report thorough simulations and several challenging real experiments using a 6 degrees-of-freedom robotic arm controlled from dual-hemispherical acquisitions.**

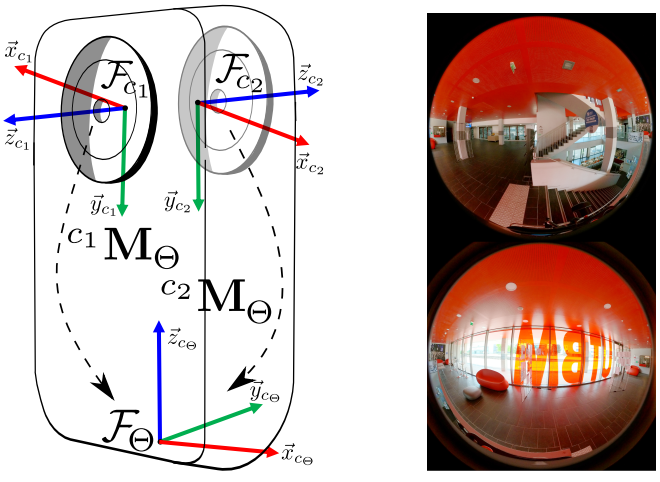
I. Introduction

A. Motivation

Unconventional cameras, such as event cameras, polarization cameras, light-field cameras to name but a few, have made significant contributions to the field of robotics by enhancing perception systems, and thus improving performance and autonomy of robots in various domains. Among these unusual imaging devices, wide-angle acquisition systems are very interesting for robot perception since capturing a large view of the surroundings allows obtaining a comprehensive awareness of the environment in a few or even a single acquisition. This is useful for numerous robotic applications where a wide field of view (FoV) increases the chances to see reliable visual information and are thus studied for various tasks such as object detection [1], obstacle avoidance [2], navigation [3], localization and mapping [4] or visual servoing (VS) [5]. The latter is a technique used to control the motion of a dynamic system, such as a robot, using visual information as feedback [6]. The robot velocities are iteratively computed in order to guide it to a specific pose in the workspace, generally by minimizing an error between current visual features and desired ones.

The VS problem has been very well studied over time. Various modeling for different types of cameras have been developed. It has been demonstrated that a wide-angle camera of an approximately hemispherical FoV offers several advantages, such as robustness to occlusions and convergence domain enlargement [7]. In parallel, various visual features have been proposed. Dense visual features, i.e., features that use the whole photometric information of the images by exploiting all their pixels, provide numerous benefits, such as a high accuracy at convergence without the need of extracting and tracking geometric features [6]. In this work, we combine the potential of those considerations, extending the entire photometric information to the whole 360-degree FoV.

There is a variety of acquisition systems designed to capture a 360-degree view. A camera mounted on a motorized pan-tilt head can successively acquire a series of different shots in order to cover an entire scene. High-quality panoramas in terms of size and resolution are thus made, but the acquisition duration is not suitable for VS. Other optical rigs use multiple synchronized cameras positioned strategically to capture a full sphere of images. Obviously, the smaller the FoV of each camera, the more cameras are needed to cover the entire scene. In our context, we are interested in systems composed of the smallest number of cameras because they are more compact, and therefore easier to attach to a robot. Thus, in this work, we use an optical rig composed of two fisheye wide-angle lenses oriented in opposite directions (Fig. 1a). An acquisition with this kind of device produces a synchronized pair of hemispherical images (Fig. 1b) that are often stitched together to build an equirectangular representation of the captured scene [8]. In our case, we directly work on the Dual-Hemispherical (DH)



(a) (b)

Fig. 1: Dual-Hemispherical camera: (a) a sketch of a 360-degree optical rig composed of two back-to-back fisheye wide-angle lenses, (b) an example of a captured image pair.

[1]Nathan Crombez and Jocelyn Buisson are with UTBM, CIAD, F-90010 Belfort, France nathan.crombez@utbm.fr, jocelyn.buisson@utbm.fr

[2]Antoine N. André and Guillaume Caron are with CNRS-AIST JRL, IRL, Tsukuba, Japan antoine.andre@aist.go.jp

[3]Guillaume Caron is also with UPJV, MIS lab, Amiens, France guillaume.caron@u-picardie.fr

acquisition in order to save processing time and work with the raw captured pixels.

B. Related works

1) Visual Servoing with a camera of hemispherical FoV: Due to their very large FoV, fisheye and catadioptric cameras have attracted interest for robot VS when large rotations are needed. Indeed, such large rotations can prevent cameras of conventional FoV to perceive the visual features corresponding to those of the desired image. Hence, VS with cameras of hemispherical FoV was formulated for features such as points [9], lines [10], [11], and spherical objects [12] that must be detected and matched between the current image and the desired one for the control law to be computed.

Perspective, catadioptric, and some types of wide-angle cameras, including hemispherical cameras having a single viewpoint, can be modeled with the Unified Central projection Model (UCM) [9]. It describes the image geometrical formation by projecting the scene on a unit sphere, then to the image plane. The UCM, thanks to the sphere of projection it involves, brings interesting control properties decoupling the degrees of freedom for point-based VS [13], [14], thus overcoming the interest of seeing wider with a hemispherical camera than a conventional one.

2) Hemispherical Photometric Visual Servoing (HPVS): The visual feature of lowest level in the VS literature is the direct use of pixel brightness as input to the control law. The seminal work of photometric VS for hemispherical camera [7] controls the robot embedding the camera in order to regulate to zero the pixel-to-pixel differences between the current and the desired images (Sum of Squared Differences of brightness, SSD). The major interests of HPVS are its high accuracy at convergence and the absence of feature detection and matching. Other ways to compare the current image brightness to the desired ones have been researched to reach a higher robustness to scene changes than with the SSD, e.g. with the Normalized Mutual Information [15]. However, the convergence domain remains tight.

More recently, transforming the images as dense Photometric Gaussian Mixtures could enlarge significantly the convergence domain of the direct alignment of hemispherical images to dense colored 3D point clouds [16] but the computation complexity currently confines that visual feature to offline alignment, thus preventing its use in real-time VS.

3) Visual Servoing with multiple cameras: The most obvious use of multiple cameras on a robot for visual servoing is to bring sensing redundancy to reconstruct the 3D coordinates of geometric features. For instance, the depth of feature points [17] or of the centroid of a segmented image region are computed from stereo-vision using from two [18] to arrays of up to nine conventional cameras mounted on the end-effector of a robot arm [19] to deal with occlusions. A combination of four static color-depth (RGB-D) sensors pointing to the robot arm end-effector workspace and a stereo camera mounted on the arm [20] has also been researched for avoiding occlusions for VS in the context of manipulation.

Of course, several cameras can also be installed without overlapping their FoV and used for visual servoing too. However, the classical use of cameras of conventional FoV combined with the need for the visual features corresponding to the desired ones to stay in each FoV, the more the conventional cameras, the tighter the convergence domain [21].

C. Contributions

The central idea of this work is to exploit a maximum of direct visual information around a robot for its VS control. To achieve this, we modeled and developed the Dual-Hemispherical Photometric Visual Servoing (DHPVS). This letter describes the following contributions:

- DH camera modeling,
- explicit formulation of DHPVS features and interaction matrix,
- experimental demonstration of the effect of the 360-degree vision on robot behavior and trajectories.

The proposed method is evaluated and compared with state-of-the-art wide-angle photometric VS on a 6 degrees-of-freedom robot arm through thorough simulations and several challenging experiments.

D. Outline

The remainder of this letter is organized as follows. Section II introduces the modeling of the DH camera. Then, Section III presents the formulation of DHPVS including the visual feature modeling III-A and the interaction matrix modeling III-B. The validation of the method in simulation is presented in Section IV. Experimental results on a real robot including comparative, qualitative and quantitative evaluations are presented in Section V. Finally, conclusions and future works are described in Section VI.

II. Dual-Hemispherical camera modeling

A. Extrinsic modeling

Each of the two wide-angle cameras has its own orthonormal frame (Fig. 1a), respectively $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$. A unique orthonormal frame $\mathcal{F}_\Theta$ is also considered for the whole DH camera. The pose of each wide-angle camera $c_i, i \in \{1,2\}$ with respect to the frame $\mathcal{F}_\Theta$ is defined by a 4×4 homogeneous matrix:

$$ {}^{c_i}\tilde{\mathbf{M}}_{\Theta(4\times4)} = \begin{bmatrix} {}^{c_i}\mathbf{R}_{\Theta(3\times3)} & {}^{c_i}\mathbf{t}_{\Theta(3\times1)} \\ \mathbf{0}_{(1\times3)} & 1 \end{bmatrix},\ i \in \{1,2\}, \quad (1) $$

where ${}^{c_i}\mathbf{R}_\Theta \in SO(3)$ is a rotation matrix and ${}^{c_i}\mathbf{t}_\Theta \in \mathbb{R}^3$ is a translation vector. Note that the notation $\tilde{\cdot}$ indicates an element expressed using homogeneous coordinates.

In the following, a pose is also represented with its minimal form, e.g., $\mathbf{r} = (\mathbf{t}, \theta\mathbf{w})$ where $\mathbf{t}$ describes the translation part, while the rotation part is expressed under the form $\theta\mathbf{w}$, where $\mathbf{w}$ represents a unit rotation-axis vector and θ a rotation angle around this axis.

An acquisition with the DH camera at pose $\mathbf{r}_\Theta$ is noted:

$$ \mathbf{\Theta}(\mathbf{r}_\Theta) = \{\mathbf{I}_{c_1}(\mathbf{r_1}), \mathbf{I}_{c_2}(\mathbf{r_2})\}, \quad (2) $$

where $\mathbf{I}_{c_1}$ and $\mathbf{I}_{c_2}$ are respectively the hemispherical images captured by the wide-angle cameras c_1 and c_2. The poses $\mathbf{r_1}$ and $\mathbf{r_2}$ are thus rigidly linked to the pose $\mathbf{r}_\Theta$ with the transformations ${}^{c_1}\tilde{\mathbf{M}}_\Theta$ and ${}^{c_2}\tilde{\mathbf{M}}_\Theta$ (Fig. 1a). Both hemispherical images $\mathbf{I}_{c_i},\ i \in \{1,2\}$ have the same size, each pixel has a location $\mathbf{u} = [u, v]^\top$ and an intensity noted $I_{c_i}(\mathbf{u})$.

B. Intrinsic modeling

We use the UCM (Sec. I-B.1) to describe independently each hemispherical camera constituting the whole DH acquisition rig. According to the unified central projection model, a 3D point ${}^{c_i}\mathbf{X} = [{}^{c_i}X, {}^{c_i}Y, {}^{c_i}Z]^\top$ expressed in the coordinates system of either the wide-angle camera c_1 or c_2, is first projected on a unit sphere $\mathcal{S}_i$, centered at the projection center c_i:

$$ {}^{c_i}\mathbf{X}_\mathcal{S} = \begin{bmatrix} {}^{c_i}X_\mathcal{S} \\ {}^{c_i}Y_\mathcal{S} \\ {}^{c_i}Z_\mathcal{S} \end{bmatrix} = \begin{bmatrix} \frac{{}^{c_i}X}{\rho} \\ \frac{{}^{c_i}Y}{\rho} \\ \frac{{}^{c_i}Z}{\rho} \end{bmatrix}, \tag{3} $$

where $\rho = \sqrt{{}^{c_i}X^2 + {}^{c_i}Y^2 + {}^{c_i}Z^2}$. The spherical point ${}^{c_i}\mathbf{X}_\mathcal{S}$ is then projected on the camera sensor plane by a perspective projection considering a distance ξ between the unit sphere center and the second perspective projection center:

$$ \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{{}^{c_i}X_\mathcal{S}}{{}^{c_i}Z_\mathcal{S}+\xi} \\ \frac{{}^{c_i}Y_\mathcal{S}}{{}^{c_i}Z_\mathcal{S}+\xi} \end{bmatrix} = \begin{bmatrix} \frac{{}^{c_i}X}{{}^{c_i}Z+\xi\rho} \\ \frac{{}^{c_i}Y}{{}^{c_i}Z+\xi\rho} \end{bmatrix}. \tag{4} $$

Finally, the image point is obtained following the classical sensor to image conversion:

$$ \tilde{\mathbf{u}} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}\tilde{\mathbf{x}}, \tag{5} $$

where α_u and α_v are respectively the horizontal and vertical generalized scale factors, and where u_0 and v_0 are the coordinates of the principal point in pixels. Therefore, in this work, we describe the whole DH camera by two separate sets of intrinsic parameters $\mathbf{k}_i = \{\alpha_{u_i}, \alpha_{v_i}, u_{0_i}, v_{0_i}, \xi_i\}, i \in \{1,2\}$ and two rigid transformations ${}^{c_i}\tilde{\mathbf{M}}_\Theta$, i.e., one for each wide-angle camera (Fig. 1a).

III. Dual-hemispherical photometric visual servoing

The core of a VS scheme is the interaction matrix $\mathbf{L_s}$ that relates the time variation of the considered visual features $\mathbf{s}$ to the camera velocity $\mathbf{v}$, similar to $\dot{\mathbf{r}}$ as defined by [6]:

$$ \dot{\mathbf{s}} = \mathbf{L_s}\mathbf{v}. \tag{6} $$

In this work, $\mathbf{v}$ is computed by iteratively minimizing the visual differences between DH acquisitions captured during the visual servoing and a desired DH acquisition captured beforehand. This section defines both the considered photometric DH visual feature and its related interaction matrix.

A. Dual-hemispherical photometric visual features

As explained before, a DH acquisition is considered as a pair of hemispherical images. In order to leverage the entire information they contain, and to avoid any image processing to extract and match sparse visual features between two pairs of wide-angle images, we propose to use the whole DH photometric information as a dense visual feature. The proposed visual feature is the stacking of every pixel intensity of both hemispherical images $\mathbf{I}_{c_1}$ and $\mathbf{I}_{c_2}$ such as:

$$ \mathbf{s} = \bar{\bar{\mathbf{\Theta}}} = \begin{bmatrix} \bar{\mathbf{I}}_{c_1} \\ \bar{\mathbf{I}}_{c_2} \end{bmatrix} = \begin{bmatrix} I_{c_1}(\mathbf{u}_0) \\ I_{c_1}(\mathbf{u}_1) \\ \vdots \\ I_{c_1}(\mathbf{u}_{W-1\times H-1}) \\ I_{c_2}(\mathbf{u}_0) \\ I_{c_2}(\mathbf{u}_1) \\ \vdots \\ I_{c_2}(\mathbf{u}_{W-1\times H-1}) \end{bmatrix}, \tag{7} $$

with W and H, respectively, the width and height of both images and where the camera poses are omitted for compactness. The overline symbol $\bar{\cdot}$ denotes a vectorization, and $\bar{\bar{\cdot}}$ denotes a stacking of vectorizations.

The error that has to be regulated to zero is then given by:

$$ \mathbf{e} = \mathbf{s}(\mathbf{r}) - \mathbf{s}^* \tag{8} $$

$$ = \bar{\bar{\mathbf{\Theta}}}(\mathbf{r}) - \bar{\bar{\mathbf{\Theta}}}^*, \tag{9} $$

where $\cdot^*$ denotes the desired set of visual features, i.e., the stacking of the vectorization of both hemispherical images acquired at the desired pose $\mathbf{r}^*$.

From equations (6) and (9), a classical Gauss-Newton scheme can be used to design the control law to regulate to zero the DH photometric error, leading to:

$$ \mathbf{v} = -\lambda \mathbf{L}_\mathbf{\Theta}^\dagger \mathbf{e}, \tag{10} $$

where $\lambda \in \mathbb{R}$ is a positive scalar that tunes the decrease rate of the error and where $\cdot^\dagger$ denotes the matrix pseudo-inverse. The modeling of the interaction matrix $\mathbf{L_\Theta}$ that links the variation of all pixel intensities contained in the overall DH acquisition $\mathbf{\Theta}$ to the camera motion is detailed hereafter.

B. Dual-hemispherical photometric interaction matrix

If we consider that c_1 and c_2 are two independent wide-angle cameras, equations (6) and (7) could lead to:

$$ \begin{bmatrix} \dot{\bar{\mathbf{I}}}_{c_1} \\ \dot{\bar{\mathbf{I}}}_{c_2} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{\mathbf{I}_{c_1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{\mathbf{I}_{c_2}} \end{bmatrix} \begin{bmatrix} {}^{c_1}\mathbf{v_1} \\ {}^{c_2}\mathbf{v_2} \end{bmatrix}, \tag{11} $$

where ${}^{c_1}\mathbf{v_1}$ and ${}^{c_2}\mathbf{v_2}$ are the velocities of respectively cameras c_1 and c_2 expressed in their own frame. The interaction matrices $\mathbf{L}_{\mathbf{I}_{c_i}}, i \in \{1,2\}$ are built from the dense photometric features contained in the images $\mathbf{I}_{c_i}$. Each of these Jacobian matrices are composed as follows [7]:

$$ \mathbf{L}_{\mathbf{I}_{c_i}} = -\nabla\mathbf{I}_{c_i}\mathbf{L}_{\mathbf{u}c_i}\mathbf{L}_{\mathbf{x}c_i}, \tag{12} $$

where:

$$ \nabla\mathbf{I}_{c_i} = \begin{bmatrix} \frac{\delta\mathbf{I}_{c_i}}{\delta\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \frac{\delta\mathbf{I}_{c_i}}{\delta u} & \frac{\delta\mathbf{I}_{c_i}}{\delta v} \end{bmatrix} \tag{13} $$

are spatial gradients approximated with finite differences (i.e., convolution with differentiation kernel), where:

$$\mathbf{L}_{\mathbf{u}c_i} = \begin{bmatrix} \frac{\delta \mathbf{u}}{\delta \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \alpha_{u_i} & 0 \\ 0 & \alpha_{v_i} \end{bmatrix}, \tag{14}$$

with α_{u_i} and α_{v_i} are the generalized scale factors defined in Section II-B, and where:

$$\begin{aligned} \mathbf{L}_{\mathbf{x}c_i} &= \begin{bmatrix} \frac{\delta \mathbf{x}}{\delta \mathbf{r}_i} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1+x^2(1-\xi(\alpha+\xi))+y^2}{\rho(\alpha+\xi)} & \frac{\xi xy}{\rho} & \frac{\alpha x}{\rho} \\ \frac{\xi xy}{\rho} & -\frac{1+y^2(1-\xi(\alpha+\xi))+x^2}{\rho(\alpha+\xi)} & \frac{\alpha y}{\rho} \\ xy & -\frac{(1+x^2)\alpha-\xi y^2}{\alpha+\xi} & y \\ \frac{(1+y^2)\alpha-\xi x^2}{\alpha+\xi} & -xy & -x \end{bmatrix} \end{aligned} \tag{15}$$

are the interaction matrices that relate the points variation in the sensor frame with respect to the wide-angle camera pose variation with $\alpha = \sqrt{1+(1+\xi^2)(x^2+y^2)}$. The points distance ρ involved in the computation of the interaction matrices is actually unknown. Practically, a single constant value is generally used for every pixel [7].

In our case, the two wide-angle cameras are not independent since they belong to a same 360-degree optical rig. Then, the camera velocities ${}^{c_1}\mathbf{v_1}$ and ${}^{c_2}\mathbf{v_2}$ can be expressed in the same reference frame such as the frame related to the whole system $\mathcal{F}_\Theta$ (Fig. 1a) leading to a single velocity vector ${}^{\Theta}\mathbf{v}$. Assuming that the DH camera is fully calibrated, it is possible to express ${}^{c_1}\mathbf{v_1}$ and ${}^{c_2}\mathbf{v_2}$ with respect to $\mathcal{F}_\Theta$:

$${}^{\Theta}\mathbf{v} = \begin{cases} {}^{\Theta}\boldsymbol{\mathcal{V}}_{c_1} \; {}^{c_1}\mathbf{v_1} \\ {}^{\Theta}\boldsymbol{\mathcal{V}}_{c_2} \; {}^{c_2}\mathbf{v_2} \end{cases}, \tag{16}$$

where ${}^{c_i}\boldsymbol{\mathcal{V}}_\Theta$ are twist transformation matrices such as:

$${}^{c_i}\boldsymbol{\mathcal{V}}_\Theta = \begin{bmatrix} {}^{c_i}\mathbf{R}_\Theta & [{}^{c_i}\mathbf{t}_\Theta]_\times \, {}^{c_i}\mathbf{R}_\Theta \\ \mathbf{0} & {}^{c_i}\mathbf{R}_\Theta \end{bmatrix}, \; i \in \{1,2\}, \tag{17}$$

in which $[{}^{c_i}\mathbf{t}_\Theta]_\times$ are the skew symmetric matrices of the translation vectors ${}^{c_i}\mathbf{t}_\Theta$.

Injecting equation (16) in equation (11) and after rearrangements, we obtain:

$$\begin{bmatrix} \dot{\mathbf{I}}_{c_1} \\ \dot{\mathbf{I}}_{c_2} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{\mathbf{I}_{c_1}} \; {}^{c_1}\boldsymbol{\mathcal{V}}_\Theta \\ \mathbf{L}_{\mathbf{I}_{c_2}} \; {}^{c_2}\boldsymbol{\mathcal{V}}_\Theta \end{bmatrix} \; {}^{\Theta}\mathbf{v}, \tag{18}$$

Consequently, the interaction matrix that relates the variation of pixel intensities of a whole DH acquisition with respect to the camera motion is:

$$\mathbf{L}_{\boldsymbol{\Theta}} = \begin{bmatrix} \mathbf{L}_{\mathbf{I}_{c_1}} \; {}^{c_1}\boldsymbol{\mathcal{V}}_\Theta \\ \mathbf{L}_{\mathbf{I}_{c_2}} \; {}^{c_2}\boldsymbol{\mathcal{V}}_\Theta \end{bmatrix}. \tag{19}$$

$\mathbf{L}_{\boldsymbol{\Theta}}$ allows computing the single ${}^{\Theta}\mathbf{v}$ the way recommended by previous works [18] that shown the independent use of images of multiple cameras to compute a single velocity vector without considering the frame change between the cameras can lead to uncontrolled motions.

IV. Evaluation in Simulation

We first validate and evaluate the proposed DHPVS in simulation. A DH camera is simulated within a synthetic environment using the game engine Unity [22]. The 3D virtual environment is a house interior made up of several rooms containing different furniture, decorations and lights (Fig. 2a). The virtual DH camera consists in the simulation of two wide angle cameras rigidly linked to a common reference frame. A wide angle image is obtained by first projecting the surroundings onto a unit sphere, and then rendering this sphere with a dedicated fragment shader. This process reproduces the unified central projection model, and thus, produces images with wide-angle distortions (Fig. 2b). Each camera of the virtual DH rig has its own configurable set of intrinsic parameters $\mathbf{k}_i$ (see Sec. II-B).

(a) (b)

Fig. 2: Simulation: (a) the 17 desired poses randomly distributed throughout the synthetic environment used in the experiments, and (b) a sample of a synthetic DH acquisition.

In order to evaluate the contribution of using the whole surrounding acquisition as visual features, we compare DHPVS to single camera Hemispherical Photometric Visual Servoing (HPVS) [7] using either, camera c_1 only (HPVS$_{c_1}$) or camera c_2 only (HPVS$_{c_2}$). Despite the availability of all the depths in simulation, we used the same depth value ($\rho = 1.0m$) for every pixel to compute the interaction matrix. This choice is made to be in the same conditions as for real experiments where the depth is unknown (Sec. V). HPVS has already proved to be robust to such approximation [7] so it is reasonable to expect DHPVS will be too, as shown in the rest of this paper. The synthetic environment is covered with 17 desired camera poses randomly distributed throughout the scene (Fig. 2a). For each of these 17 desired poses $p \in \{1, 2, ..., 17\}$, three initial poses are randomly generated with an increased degree of difficulty. More precisely, for the first degree of difficulty d_1, the initial displacement $\Delta\mathbf{r}_{p,1} = (\Delta\mathbf{t}_{p,1}, \Delta\theta\mathbf{w}_{p,1})$ is set as $\Delta\mathbf{t}_{p,1} \in \pm[0m, 0.33m]^3$ and $\Delta\theta\mathbf{w}_{p,1} \in \pm[0°, 7.5°]^3$; for the second one d_2: $\Delta\mathbf{t}_{p,2} \in \pm[0.33m, 0.66m]^3$ and $\Delta\theta\mathbf{w}_{p,2} \in \pm[7.5°, 15°]^3$; and for the last d_3: $\Delta\mathbf{t}_{p,3} \in \pm[0.66m, 1.0m]^3$ and $\Delta\theta\mathbf{w}_{p,3} \in \pm[15°, 22.5°]^3$. For every VS run following this protocol, four metrics are computed:

- Convergence Success (CS): we consider that an experiment has converged if the error in position is less than $1.0cm$ and the error in orientation is less than $1.0°$,
- Trajectory Length (TL): total distance travelled by the camera,
- Trajectory Area (TA): area between the camera trajectory and the straight line joining the initial and desired positions.
- Condition Number (CN): conditioning of the interaction matrix.

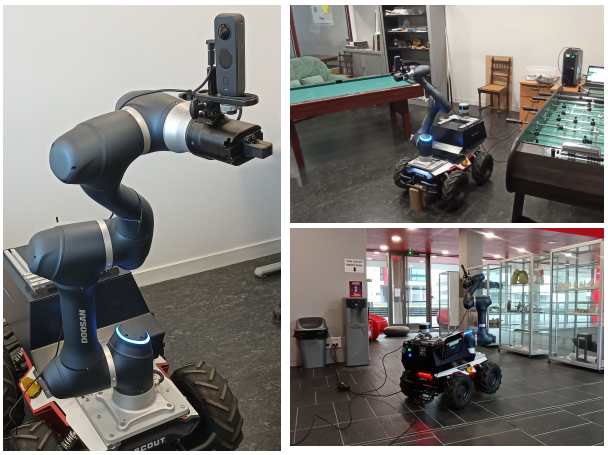
The averages of these metrics for each method (DHPVS, HPVS_{c_1}, HPVS_{c_2}) with respect to the degree of difficulty (d_1, d_2, d_3) are reported in TABLE I. DHPVS' higher success rate than HPVS', regardless of the degree of difficulty, shows that DHPVS is less inclined to diverge, or to fall into a local minimum, than when a single hemispherical camera is used. The trajectories obtained with DHPVS are shorter than those of HPVS, indicating DHPVS produces velocities that allow the camera to converge to the desired pose with straighter trajectories than HPVS does. The TA metric can be seen as a quantification of the trajectory scale. We can therefore deduce that, in addition to the above-mentioned qualities, DHPVS produces narrower and less curved camera trajectories than HPVS. It is interesting to note that the conditioning is always better when both hemispheres are used, than with one of them only. An ill-conditioned interaction matrix can have various significant impacts on the VS behaviors, e.g., instability, slow convergence, limited workspace or sensitivity and robustness issues [23]. Since the condition number of a VS interaction matrix indicates a global measure of the motion visibility, having the widest possible FoV and working with the whole DH content is very advantageous, especially with regard to the camera trajectory.

	d_1				d_2				d_3			
	$\overline{\text{CS}}$	$\overline{\text{TL}}$	$\overline{\text{TA}}$	$\overline{\text{CN}}$	$\overline{\text{CS}}$	$\overline{\text{TL}}$	$\overline{\text{TA}}$	$\overline{\text{CN}}$	$\overline{\text{CS}}$	$\overline{\text{TL}}$	$\overline{\text{TA}}$	$\overline{\text{CN}}$
DHPVS	**100**	**0.41**	**1.12**	**2.92**	**88**	**1.43**	**4.50**	**3.04**	**71**	**2.44**	**7.30**	**2.54**
HPVS_{c_1}	88	0.64	1.37	5.12	71	2.29	5.10	4.83	65	4.45	27.19	5.19
HPVS_{c_2}	**100**	0.66	1.71	5.34	71	2.39	5.85	4.90	35	4.07	8.21	4.99

TABLE I: Simulation evaluation results: Averages of Convergence Success (CS) in percent, Trajectory Length (TL) in centimeters, Trajectory Area (TA) in square centimeters and Condition Number (CN), regarding the 3 degrees of difficulty considered (d_1, d_2, d_3). The bold font indicates the best results.

V. Experimental results

Real experiments are performed using a 6-axis industrial robot (Doosan A0509) with a DH camera (Insta360 ONE X2) mounted on its end-effector (Fig.3a) within various environments (Fig.3b). The DH camera is made of a front fisheye camera and a back one of the same FoV of 191°, meaning the two cameras overlap by approximately 22° degrees. The DH camera has been intrinsically and extrinsically calibrated [24]. Thus, we know for both wide-angle cameras their respective sets of intrinsic parameters and their poses with respect to the robot end-effector. The latter are, in their minimal form (see Sec. II-A): $\mathbf{r}_{c_1} = (0.162m, -0.015m, 0.066m, 71.1°, -71.1°, 66.5°)$ for the first fisheye of the DH camera and $\mathbf{r}_{c_2} = (0.158m, 0.004m, 0.064m, -66.5°, 68.8°, 70.5°)$ for the second one. Without loss of generality, we set $\mathcal{F}_\Phi$ at the robot end-effector frame. DHPVS has been implemented using the ROS middleware [25] and the ViSP library [26]. All our developments and implementations are publicly available at https://github.com/NathanCrombez/DHPVS.



(a) (b)

Fig. 3: Experimental setup: (a) Doosan A0509 6-axis robot arm with an Insta360 ONE X2 DH camera on its end-effector, and (b) different environments where experiments were carried out (the mobile base is not used in this work).

A. Comparison with HPVS

The aim of these first real experiments is to verify the observations arising from the comparison of DHPVS with HPVS in simulation (Section IV).

1) Real experiment #1: The first real experiment involves a deliberately short displacement between the initial and desired camera poses in order to compare the behavior of DHPVS with both HPVS on a simple case. More precisely, the initial displacement is: $\Delta\mathbf{r}_i = (0.076m, 0.019m, 0.001m, 8.1°, -1.4°, 9.3°)$. The initial difference in image space, i.e., the difference between the desired DH acquisition (Fig. 4a) and the initial one (Fig. 4b), is consequently not too important (Fig. 4c). DHPVS converges to the desired pose following almost a straight camera trajectory (Fig. 4f). The final pose error is very small: $\Delta\mathbf{r}_f = (0.0mm, -0.1mm, -0.0mm, 0.0°, 0.0°, 0.0°)$, as well as the final image of differences (Fig. 4d). HPVS using either camera c_1 or camera c_2 converges also accurately to the desired pose. However, despite the small displacement that had to be corrected, HPVS trajectories are significantly more curvy than DHPVS ones (Fig. 4f).

2) Real experiment #2: The second experiment is conducted to appreciate the value of using the whole photometric information available all around the scene when a large motion from the initial pose to the desired one has to be performed. More precisely, the initial displacement is: $\Delta\mathbf{r}_i = (0.499m, -0.185m, 0.326m, -2.4°, -29.5°, 13.6°)$. Because of this significant displacement, the initial photometric error (square norm of (9)) between the desired and initial DH acquisitions (Fig. 5a and Fig. 5b) increases by 40% compared to Real experiment #1 (see Fig. 4e versus Fig. 5e). Despite these challenging conditions, DHPVS succeeds to control the camera until the desired pose is reached. As shown by the final image of differences (Fig. 5d), the convergence accuracy is very good, and thus the final pose error is very small: $\Delta\mathbf{r}_f = (0.3mm, 0.0mm, -0.2mm, -0.0°, -0.0°, -0.0°)$. On the other hand, HPVS using either camera c_1 or camera c_2 failed to control the robot to bring the camera to the desired pose. Indeed, due to the excessively wide trajectories that the camera achieved (Fig. 5f), the robotic arm reached the limits of its workspace during both HPVS.

3) Real experiment #3: The third experiment involves an even greater displacement, particularly in terms of rotation around the three axes. Because of this significant displacement, the photometric error between the desired and initial DH acquisitions (Fig. 6a and Fig. 6b) increases by 15% compared to Real experiment #2 (see Fig. 5e versus Fig. 6e). The displacement between the desired pose and the initial one is: $\Delta\mathbf{r}_i = (0.395m, -0.221m, 0.253m, 11.6°, -31.0°, -14.9°)$. Even if the initial DH difference is important (Fig. 6c),

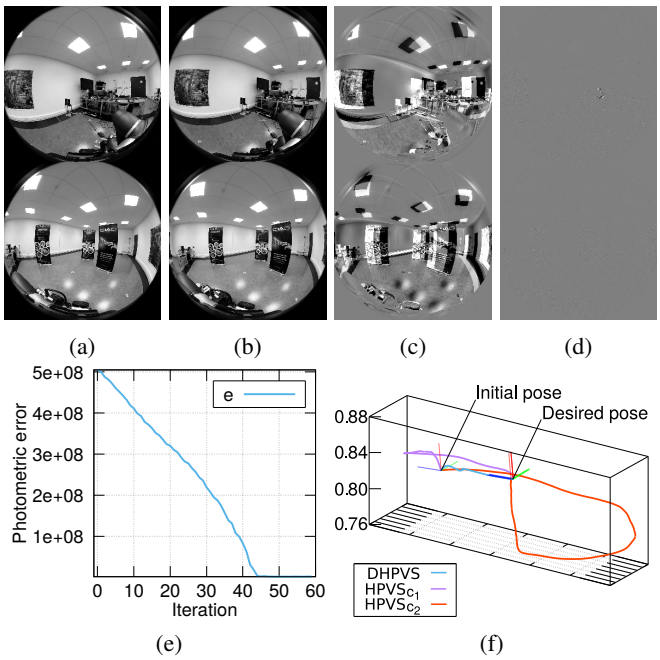


Fig. 4: Real experiment #1: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences, (e) DH photometric error and (f) trajectories.

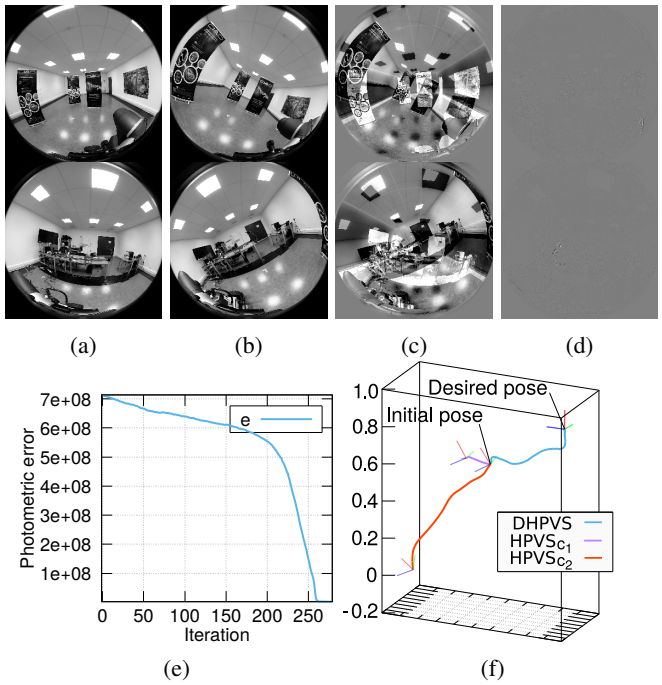


Fig. 5: Real experiment #2: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences, (e) DH photometric error and (f) trajectories.

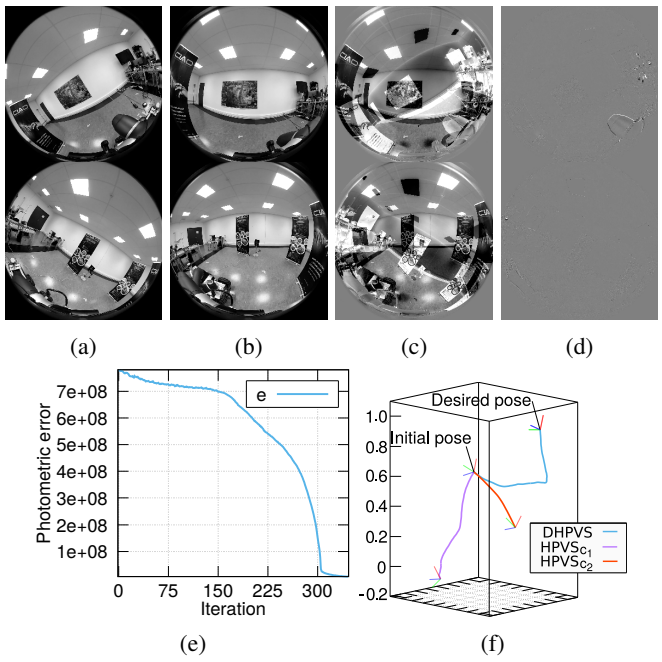


Fig. 6: Real experiment #3: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences, (e) DH photometric error and (f) trajectories.

DHPVS successfully controlled the robotic arm in order to drive the camera to the desired pose. As it can be seen on the final acquisitions difference (Fig. 6d), the convergence accuracy is very good, confirmed by the final pose error: $\Delta\mathbf{r}_f = (0.3mm, -1.9mm, 0.1mm, -0.0°, 0.0°, 0.0°)$. On

the contrary, HPVS using either camera c_1 or camera c_2 failed due to excessively sinuous or wide trajectories (Fig. 6f). More specifically, on the one hand, the velocities computed with HPVS_{c_1} caused the arm to retract, bringing it into a self-collision posture. On the other hand, the broad trajectory obtained with HPVS_{c_2} brought the robotic arm to the limits of its workspace. These experimental results confirm the conclusions drawn from the simulation evaluations (Sec. IV). Indeed, using the whole photometric information around the camera leads to straighter trajectories than using the half of it, enabling the robot to correct larger displacements while avoiding reaching its workspace bounds.

4) Real experiment #4: This experiment is intended to challenge DHPVS beyond its limits. The displacement between the desired pose and the initial one is: $\Delta\mathbf{r}_i = (0.115m, 0.145m, -0.138m, 4.6°, 51.1°, 41.8°)$. Due to the large initial orientation gap (80% larger than for Real experiment #3, particularly three times larger around the optical axis), the visual content of the initial DH acquisition (Fig. 7b) and the desired one (Fig. 7a) are very misaligned (Fig. 7c). Under such conditions, neither DHPVS nor HPVS succeeded in bringing the camera to the desired pose. They fall into local minima (Fig. 7d). This is not surprising because no direct visual servoing method could reach global convergence yet, though DHPVS has practically shown a larger convergence domain than HPVS.

B. Robustness to perturbations

The aim of this series of real experiments is to show how DHPVS behaves in response to various perturbations in different environments.

1) Dynamic occlusions: This experiment is to study the behavior of DHPVS while operating in a dynamic environment (Fig. 3b). Four persons were moving freely around the robot throughout the experiment. Since the persons were always moving, their postures were never the same in either the desired DH acquisition (Fig. 8a) or current ones (Fig. 8b). The displacement between the desired pose and the initial one is: $\Delta\mathbf{r}_i = (0.179m, -0.397m, 0.366m, -2.8°, -33.5°, -34.4°)$.

(a) (b) (c) (d)

Fig. 7: Real experiment #4: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences.

Despite the dynamic occlusions, DHPVS still converged and brought the camera to the desired pose. Obviously, the final visual error is not completely uniform, since the scene has been continuously altered (Fig. 8d). Nevertheless, these dynamic occlusions have not affected significantly the final pose error that is very small: $\Delta\mathbf{r}_f = (-0.4mm, 2.4mm, 0.4mm, 0.0°, -0.1°, -0.1°)$.

To compare, the same experiment was carried out with the environment remaining static. Fig. 8e compares the photometric errors e_{dyn} and e_{sta} for the dynamic and static experiments, respectively. It is interesting to note that the curves follow a broadly similar progression, except that e_{dyn} is higher and noisier due to the continuous changes in the scene. The trajectories of the camera for both experiments are also quite similar (Fig. 8f), showing that the perturbations had only a minor impact on DHPVS.

2) Specularity and transparency: The aim of this experiment is to study the behavior of DHPVS while operating in a complex real-world scene that contain many reflective and transparent objects (Fig. 3b). The scene's depth range is large, approximately $[1.0m, 20.0m]$, as it can be seen in the desired DH acquisition (Fig. 9a) and in the initial one (Fig. 9b). The displacement between the desired pose and the initial one is: $\Delta\mathbf{r}_i = (0.137m, 0.208m, -0.215m, 50.8°, 38.7°, -7.9°)$. Despite the scene is not Lambertian, the large displacement and the high difference between the initial and the desired DH acquisitions (Fig. 9c), DHPVS has successfully computed the velocities to control the camera motion in order to precisely reach the desired pose:

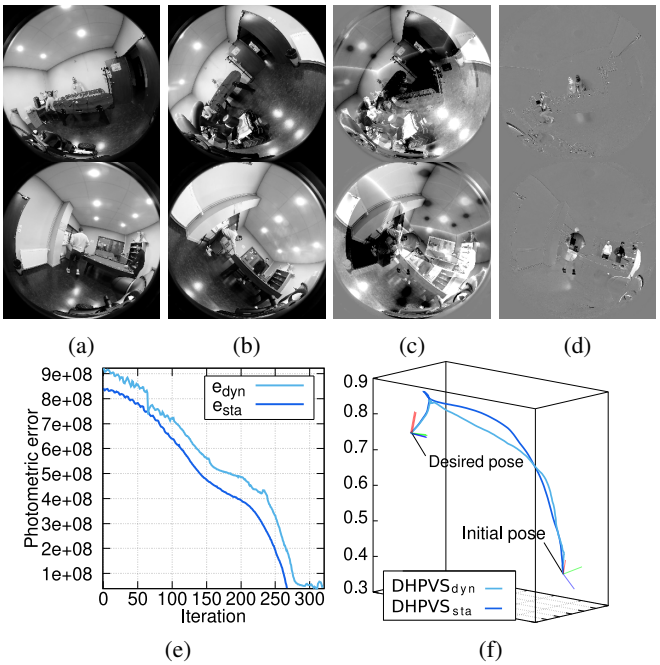


Fig. 8: Real experiment #5: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences, (e) DH photometric errors and (f) trajectories.

$\Delta \mathbf{r}_f = (0.1mm, -1.3mm, -0.1mm, 0.0°, 0.0°, 0.0°)$. We can also note that during the experiment, the ceiling lights were automatically switched off and then on twice. These perturbations can be clearly seen on the photometric error curve (Fig. 9e) around iterations 120 and 260. However, it had no impact on the behavior of DHPVS, as the trajectory shows (Fig. 9f).

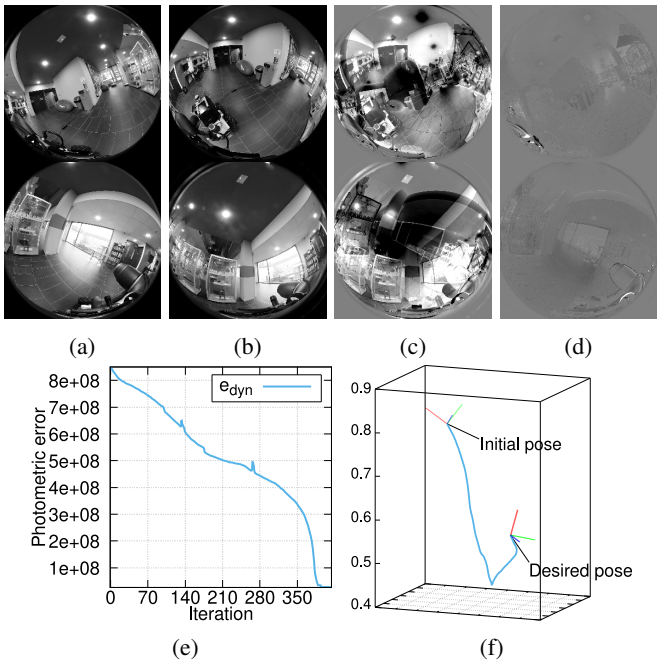


Fig. 9: Real experiment #6: (a-b) desired and initial DH acquisitions, (c-d) initial and final visual differences, (e) DH photometric error and (f) trajectory.

VI. Conclusion and Future Works

This letter introduced a visual servoing that exploit the whole surrounding visual information captured with a dual-hemispherical camera in order to control a robot. The modeling of the dual-hemispherical camera, the dual-hemispherical photometric visual feature, as well as the associated interaction matrix are detailed. Combining the photometric visual features and 360-degree acquisitions has proven to allow convergence from way farther initial errors than using a 180-degree camera with straighter camera trajectories while keeping the excellent accuracy at convergence.

Future works will extend the dual-hemispherical based visual servoing to other dense visual features, such as photometric Gaussian mixtures, and to control the whole mobile manipulator.

References

[1] H. Rashed, E. Mohamed, G. Sistu, V. R. Kumar, C. Eising, A. El-Sallab, and S. Yogamani, "Generalized object detection on fisheye cameras for autonomous driving: Dataset, representations and baseline," in *IEEE/CVF Winter Conf. on Applications of Computer Vision*, 2021, pp. 2272–2280.

[2] P. Gohl, D. Honegger, S. Omari, M. Achtelik, M. Pollefeys, and R. Siegwart, "Omnidirectional visual obstacle detection using embedded fpga," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2015, pp. 3938–3943.

[3] C.-O. Artizzu, G. Allibert, and C. Demonceaux, "Deep reinforcement learning with omnidirectional images: application to uav navigation in forests," in *Int. Conf. on Control, Autom., Robotics and Vision*, 2022.

[4] S. Ji, Z. Qin, J. Shan, and M. Lu, "Panoramic slam from a multiple fisheye camera rig," *ISPRS J. of Photogrammetry and Remote Sensing*, vol. 159, pp. 169–183, 2020.

[5] J. Ducrocq, G. Caron, and E. M. Mouaddib, "Hdromni: Optical extension of dynamic range for panoramic robot vision," *IEEE Robotics and Autom. Letters*, vol. 6, no. 2, pp. 3561–3568, 2021.

[6] F. Chaumette and S. Hutchinson, "Visual servo control, Part I: Basic approaches," *IEEE Robotics and Autom. Mag.*, vol. 13, no. 4, pp. 82–90, 2006.

[7] G. Caron, E. Marchand, and E. Mouaddib, "Photometric visual servoing for omnidirectional cameras," *Autonomous Robots*, vol. 35, 10 2013.

[8] I.-C. Lo, K.-T. Shih, and H. H. Chen, "Efficient and accurate stitching for 360° dual-fisheye images and videos," *IEEE Trans. on Image Processing*, vol. 31, pp. 251–262, 2022.

[9] J. P. Barreto, F. Martin, and R. Horaud, "Visual servoing/tracking using central catadioptric images," in *Experimental Robotics VIII*. Springer, 2003, pp. 245–254.

[10] H. Hadj-Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet, "Omnidirectional visual servoing from polar lines," in *IEEE Int. Conf. on Robotics and Autom.*, 2006, pp. 2385–2390.

[11] R. Marie, H. B. Said, J. Stéphant, and O. Labbani-Igbida, "Visual servoing on the generalized voronoi diagram using an omnidirectional camera," *J. of Intelligent & Robotic Systems*, vol. 94, pp. 793–804, 2019.

[12] R. T. Fomena and F. Chaumette, "Improvements on visual servoing from spherical targets using a spherical projection model," *IEEE Trans. on Robotics*, vol. 25, no. 4, pp. 874–886, 2009.

[13] H. Hadj-Abdelkader, Y. Mezouar, and P. Martinet, "Decoupled visual servoing from a set of points imaged by an omnidirectional camera," in *IEEE Int. Conf. on Robotics and Autom.*, 2007, pp. 1697–1702.

[14] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke, "Decoupled image-based visual servoing for cameras obeying the unified projection model," *IEEE Trans. on Robotics*, vol. 26, no. 4, pp. 684–697, 2010.

[15] B. Delabarre, G. Caron, and E. Marchand, "Omnidirectional visual servoing using the normalized mutual information," *IFAC Proceedings Volumes*, vol. 45, no. 22, pp. 102–107, 2012.

[16] S.-E. Guerbas, N. Crombez, G. Caron, and E. M. Mouaddib, "Photometric gaussian mixtures for direct virtual visual servoing of omnidirectional camera," in *IEEE CVPR Workshop on 3D Vision and Robotics*, 2021.

[17] N. Maru, H. Kase, S. Yamada, A. Nishikawa, and F. Miyazaki, "Manipulator control by using servoing with the stereo vision," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 3, 1993, pp. 1866–1870.

[18] P. Martinet and E. Cervera, "Stacking jacobians properly in stereo visual servoing," in *IEEE Int. Conf. on Robotics and Autom.*, vol. 1, 2001, pp. 717–722.

[19] C. Lehnert, D. Tsai, A. Eriksson, and C. McCool, "3d move to see: Multi-perspective visual servoing towards the next best view within unstructured and occluded environments," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2019, pp. 3890–3897.

[20] H. Cuevas-Velasquez, N. Li, R. Tylecek, M. Saval-Calvo, and R. B. Fisher, "Hybrid multi-camera visual servoing to moving target," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2018, pp. 1132–1137.

[21] E. Malis, G. Morel, and F. Chaumette, "Robot control using disparate multiple sensors," *The Int. J. of Robotics Research*, vol. 20, no. 5, pp. 364–377, 2001.

[22] Unity Technologies. Unity. [Online]. Available: https://unity.com/

[23] P. Corke and S. A. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Trans. on Robotics and Autom.*, vol. 17, pp. 507–515, 2001.

[24] R. Y. Tsai, R. K. Lenz, *et al.*, "A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration," *IEEE Trans. on Robotics and Autom.*, vol. 5, no. 3, pp. 345–358, 1989.

[25] M. Quigley, "Ros: an open-source robot operating system," in *IEEE Int. Conf. on Robotics and Autom.*, 2009.

[26] E. Marchand, F. Spindler, and F. Chaumette, "ViSP for visual servoing: a generic software platform with a wide class of robot control skills," *IEEE Robotics and Autom. Mag.*, vol. 12, no. 4, pp. 40–52, 2005.