



**HAL**  
open science

## A Direct Dense Visual Servoing Approach using Photometric Moments

Manikandan Bakthavatchalam, Omar Tahri, François Chaumette

► **To cite this version:**

Manikandan Bakthavatchalam, Omar Tahri, François Chaumette. A Direct Dense Visual Servoing Approach using Photometric Moments. *IEEE Transactions on Robotics*, 2018, 34 (5), pp.1226-1239. 10.1109/TRO.2018.2830379 . hal-01758205

**HAL Id: hal-01758205**

**<https://inria.hal.science/hal-01758205>**

Submitted on 4 Apr 2018

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

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

# A Direct Dense Visual Servoing Approach using Photometric Moments

Manikandan Bakthavatchalam, Omar Tahri and François Chaumette

**Abstract**—In this paper, visual servoing based on photometric moments is advocated. A direct approach is chosen by which the extraction of geometric primitives, visual tracking and image matching steps of a conventional visual servoing pipeline can be bypassed. A vital challenge in photometric methods is the change in the image resulting from the appearance and disappearance of portions of the scene from the camera field of view during the servo. To tackle this issue, a general model for the photometric moments enhanced with spatial weighting is proposed. The interaction matrix for these spatially weighted photometric moments is derived in analytical form. The correctness of the modelling, effectiveness of the proposed strategy in handling the exogenous regions and improved convergence domain are demonstrated with a combination of simulation and experimental results.

**Index Terms**—image moments, photometric moments, dense visual servoing, intensity-based visual servoing

## I. INTRODUCTION

Visual servoing (VS) refers to a wide spectrum of closed-loop techniques for the control of actuated systems with visual feedback [3]. A task function is defined from a set of selected visual features, based on the currently acquired image  $\mathbf{I}(t)$  and the reference image  $\mathbf{I}^*$  learnt from the desired robot pose. In a typical VS pipeline, the image stream is subjected to an ensemble of measurement processes, including one or more image processing, image matching and visual tracking steps, from which the visual features are determined. Based on the nature of the visual features used in the control law, VS methods can be broadly classified into geometric and photometric approaches. The earliest geometric approaches employ as visual features parameters observed in the image of geometric primitives (points, straight lines, ellipses, cylinders) [4]. These approaches are termed Image-based Visual Servoing (IBVS). In Pose-based Visual Servoing [5], geometric primitives are used to reconstruct the camera pose which is then used as input for visual servoing. These approaches are thus dependent on the reliable detection, extraction and subsequent tracking of the aforesaid primitives. While PBVS may be affected by instabilities in pose estimation, IBVS designed from image points may be subject to local minima, singularity, inadequate robot trajectory and limited convergence domain, when the six degrees of freedom are controlled and when the image error is large and/or when the robot has a large displacement to achieve to reach the desired pose [3]. This is due to the

strong non linearities and coupling in the interaction matrix of image points. To handle these issues, geometric moments were introduced for VS in [6]–[8], which allowed obtaining a large convergence domain and adequate robot trajectories, thanks to the reduction of the non linearities and coupling in the interaction matrix of adequate combinations of moments. However, these methods are afflicted by a serious restriction: their dependency on the availability of well-segmented regions or a set of tracked and matched points in the image. Breaking this traditional dependency, the approach proposed in this paper embraces a more general class, known as *dense VS*, in which the extraction, tracking and matching of set of points or well-segmented regions is not necessary.

In another suite of geometric methods, an homography and a projective homography are respectively used as visual features in [9] and [10], [11]. These quantities are estimated by solving a geometric or photo-geometric image registration problem, carried out with non-linear iterative methods. However, these methods require a perfect matching of the template considered in the initial and desired images, which strongly limits their practical relevance.

The second type of methods adopted the photometric approach by avoiding explicit geometric extraction and resorting instead to use the image intensities. A learning-based approach was proposed in [12], where the intensities were transformed using Principal Component Analysis to a reduced dimensional subspace. But it is prohibitive to scale this approach to multiple degrees of freedom [13]. The set of intensities in the image were directly used as visual features in [14] but the high non-linearity between the feature space and the state space limits the convergence domain of this method and does not allow obtaining adequate robot trajectories. This direct approach was later extended to omnidirectional cameras [15] and to depth map [16].

In this work, instead of using directly the raw luminance of all the pixels, we investigate the usage of visual features based on photometric moments. As it has been shown in [7] that considering geometric moments (built from a set of image points) provides a better behavior than considering directly a set of image points, we will show that considering photometric moments (built from the luminance of the pixels in the image) provides a better behavior than considering directly the luminance of the pixels. These moments are a specific case of the Kernel-based formulation in [17] which synthesized controllers only for 3D translations and rotation around the optic axis. Furthermore, the analytical form of the interaction matrix of the features proposed in [17] has not been determined, which makes impossible the theoretical sta-

Parts of this work have been presented in [1] and [2].

Manikandan Bakthavatchalam and François Chaumette are with Inria, Univ Rennes, CNRS, IRISA, Rennes, France. e-mail: Manikandan.Bakthavatchalam@inria.fr, Francois.Chaumette@inria.fr

Omar Tahri is with INSA Centre Val de Loire, Université d’Orléans, PRISME EA 2249, Bourges, France. email: omar.tahri@insa-cvl.fr

bility analysis of the corresponding control scheme. Different from [17], the interaction matrix is developed in closed-form in this paper, and most importantly taking into account all the six degrees of freedom, which is the first main contribution of this work. It is shown that this is more general as well as consistent with the current state-of-the-art.

Furthermore, an important practical (and theoretical) issue that affects photometric methods stem from the changes in the image due to the appearance of new portions of the scene or the disappearance of previously viewed portions from the camera field-of-view (FOV). This means that the set of measurements varies along the robot trajectory, with a potential large discrepancy between the initial and desired images, leading to an inconsistency between the set of luminances  $\mathbf{I}(t)$  in the current image and the set  $\mathbf{I}^*$  in the desired image, and thus also for the photometric moments computed in the current and desired images. In practice, such unmodelled disturbances influence the system behaviour and may result in failure of the control law. Another original contribution of this work is an effective solution proposed to this challenging problem by means of a spatial weighting scheme. In particular, we determine a weighting function so that a closed-form expression of the interaction matrix can be determined.

The main contributions of this paper lie in the modelling issues related to considering photometric moments as inputs of visual servoing and in the study of the improvements it brings with respect to the pure luminance method. The control scheme we have used to validate these contributions is a classical and basic kinematic controller [3]. Let us note that more advanced control schemes, such as dynamic controllers [18]–[20], could be designed from these new visual features.

The sequel of the paper is organized as follows: in Section II, the modelling aspects of photometric moments and the associated weighting strategy are discussed in depth. In Section III, the visual features adopted and the control aspects are discussed. Sections IV and V are devoted to simulations and experimental results. Finally, the conclusions drawn are presented in Section VI.

## II. MODELLING

Generalizing the classical definition of image moments, we define a weighted photometric moment of order  $(p + q)$  as:

$$m_{pq} = \iint_{\pi} x^p y^q w(\mathbf{x}) I(\mathbf{x}, t) dx dy \quad (1)$$

where  $\mathbf{x} = (x, y)$  is a spatial point on the image plane  $\pi$  where the intensity  $I(\mathbf{x}, t)$  is measured at time  $t$  and  $w(\mathbf{x})$  is a weight attributed to that measurement. By linking the variations of these moments to the camera velocity  $\mathbf{v}_c$ , the interaction matrix of the photometric moments can be obtained.

$$\dot{m}_{pq} = \mathbf{L}_{m_{pq}} \mathbf{v}_c \quad (2)$$

where  $\mathbf{L}_{m_{pq}} = [L_{m_{pq}}^{v_x} \ L_{m_{pq}}^{v_y} \ L_{m_{pq}}^{v_z} \ L_{m_{pq}}^{\omega_x} \ L_{m_{pq}}^{\omega_y} \ L_{m_{pq}}^{\omega_z}]$ . Each  $L_{m_{pq}}^{v/\omega} \in \mathbb{R}$  is a scalar with the superscripted  $v$  denoting translational velocity and  $\omega$  the rotational velocity along or around the axis  $x$ ,  $y$  or  $z$  axis of the camera frame.

Taking the derivative of the photometric moments in (1), we have

$$\dot{m}_{pq} = \iint_{\pi} x^p y^q w(\mathbf{x}) \dot{I}(x, y) dx dy \quad (3)$$

The first step is thus to model the variations in the intensity  $\dot{I}(x, y)$  that appear in (3). In [14] which aimed to use raw luminance directly as visual feature, the intensity variations were modelled using the Phong illumination model [21] resulting in an interaction matrix with parts corresponding to the ambient and diffuse terms. In practice, use of light reflection models requires cumbersome measurements for correct instantiation of the models. Besides, a perfect model should take into account the type of light source, attenuation model and different possible configurations between the light sources, the vision sensor and the target object used in the scene. Since VS is robust to modelling errors, adding such premature complexity to the models can be avoided. Instead, this paper adopts a simpler and more practical approach by using the classical brightness constancy assumption [22] to model the intensity variations, as done in [23]. This assumption considers that the intensity of a moving point  $\mathbf{x} = (x, y)$  remains unchanged between successively acquired images. This is encapsulated in the following well-known equation

$$I(\mathbf{x} + \delta\mathbf{x}, t + \delta t) = I(\mathbf{x}, t) \quad (4)$$

where  $\delta\mathbf{x}$  is the infinitesimal displacement undergone by the image point after an infinitesimal increment in time  $\delta t$ . A first order Taylor expansion of (4) around  $\mathbf{x}$  leads to

$$\nabla \mathbf{I}^T \dot{\mathbf{x}} + \dot{I} = 0 \quad (5)$$

known as the classical optic flow constraint equation (OFCE), where  $\nabla \mathbf{I}^T = [\frac{\partial I}{\partial x} \ \frac{\partial I}{\partial y}] = [I_x \ I_y]$  is the spatial gradient at the image point  $\mathbf{x}$ . Further, the relationship linking the variations in the coordinates of a point in the image with the spatial motions of a camera is well established [3]:  $\dot{\mathbf{x}} = \mathbf{L}_x \mathbf{v}_c$  where

$$\mathbf{L}_x = \begin{bmatrix} \frac{-1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & \frac{-1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix} \quad (6)$$

In general, the depth of the scene points can be considered as a polynomial surface expressed as a function of the image point coordinates [6].

$$\frac{1}{Z} = \sum_{p \geq 0, q \geq 0, p+q \leq n} A_{pq} x^p y^q \quad (7)$$

where  $n$  is the degree of the polynomial with  $n = 1$  for a planar scene. Equation (7) is a general form with the only assumption that the depth is continuous. In this work however, for simplifying the analytical forms presented, only planar scenes have been considered in the modelling<sup>1</sup>. We will see in Section V-D that this simplification is not crucial by considering non planar environments. Therefore, with  $n = 1$ , (7) becomes

$$\frac{1}{Z} = Ax + By + C \quad (8)$$

<sup>1</sup>Note that the general analytic form of  $\mathbf{L}_{m_{pq}}$  could be obtained with  $n > 1$  for non planar scenes, as was done in [6] for the geometric moments

where  $A(= A_{10}), B(= A_{01}), C(= A_{00})$  are scalar parameters that describe the configuration of the plane in the camera frame. From (5), we can write:

$$\dot{I}(x, y) = -\nabla \mathbf{I}^\top \dot{\mathbf{x}} \quad (9)$$

By plugging (8) and (6) in (9), we obtain

$$\dot{I}(x, y) = \mathbf{L}_I \mathbf{v}_c = -\nabla \mathbf{I}^\top \mathbf{L}_x \mathbf{v}_c \quad (10)$$

where  $\mathbf{L}_I = -\nabla \mathbf{I}^\top \mathbf{L}_x$  is given by:

$$\mathbf{L}_I^\top = \begin{bmatrix} I_x(Ax + By + C) \\ I_y(Ax + By + C) \\ (-xI_x - yI_y)(Ax + By + C) \\ -xyI_x - (1 + y^2)I_y \\ (1 + x^2)I_x + xyI_y \\ -yI_x + xI_y \end{bmatrix} \quad (11)$$

Substituting (10) into (3), we see that

$$\dot{m}_{pq} = \iint_{\pi} x^p y^q w(\mathbf{x}) \mathbf{L}_I \mathbf{v}_c dx dy \quad (12)$$

By comparing with (2), we can then identify and write down the interaction matrix of the photometric moments as

$$\mathbf{L}_{m_{pq}} = \iint_{\pi} x^p y^q w(\mathbf{x}) \mathbf{L}_I dx dy \quad (13)$$

Direct substitution of (11) into the above equation gives us

$$\begin{aligned} L_{m_{pq}}^{v_x} &= \iint_{\pi} x^p y^q w(\mathbf{x}) I_x (Ax + By + C) dx dy \\ L_{m_{pq}}^{v_y} &= \iint_{\pi} x^p y^q w(\mathbf{x}) I_y (Ax + By + C) dx dy \\ L_{m_{pq}}^{v_z} &= \iint_{\pi} x^p y^q w(\mathbf{x}) (-xI_x - yI_y) (Ax + By + C) dx dy \\ L_{m_{pq}}^{\omega_x} &= \iint_{\pi} x^p y^q w(\mathbf{x}) (-xyI_x - (1 + y^2)I_y) dx dy \\ L_{m_{pq}}^{\omega_y} &= \iint_{\pi} x^p y^q w(\mathbf{x}) ((1 + x^2)I_x + xyI_y) dx dy \\ L_{m_{pq}}^{\omega_z} &= \iint_{\pi} x^p y^q w(\mathbf{x}) (xI_y - yI_x) dx dy \end{aligned}$$

We see that the interaction matrix consists of a set of integro-differential equations. For convenience and fluidity in the ensuing developments, the following compact notation is introduced.

$$m_{pq}^{\nabla x} = \iint_{\pi} x^p y^q w(\mathbf{x}) I_x dx dy \quad (14a)$$

$$m_{pq}^{\nabla y} = \iint_{\pi} x^p y^q w(\mathbf{x}) I_y dx dy \quad (14b)$$

Each component of the interaction matrix in (13) can be easily re-arranged and expressed in terms of the above compact notation as follows:

$$\begin{aligned} L_{m_{pq}}^{v_x} &= A m_{p+1,q}^{\nabla x} + B m_{p,q+1}^{\nabla x} + C m_{p,q}^{\nabla x} \\ L_{m_{pq}}^{v_y} &= A m_{p+1,q}^{\nabla y} + B m_{p,q+1}^{\nabla y} + C m_{p,q}^{\nabla y} \\ L_{m_{pq}}^{v_z} &= -A m_{p+2,q}^{\nabla x} - B m_{p+1,q+1}^{\nabla x} - C m_{p+1,q}^{\nabla x} \\ &\quad - A m_{p+1,q+1}^{\nabla y} - B m_{p,q+2}^{\nabla y} - C m_{p,q+1}^{\nabla y} \\ L_{m_{pq}}^{\omega_x} &= -m_{p+1,q+1}^{\nabla x} - m_{p,q}^{\nabla y} - m_{p,q+2}^{\nabla y} \\ L_{m_{pq}}^{\omega_y} &= m_{p,q}^{\nabla x} + m_{p+2,q}^{\nabla x} + m_{p+1,q+1}^{\nabla y} \\ L_{m_{pq}}^{\omega_z} &= -m_{p,q+1}^{\nabla x} + m_{p+1,q}^{\nabla y} \end{aligned} \quad (15)$$

The terms  $m_{pq}^{\nabla x}$  and  $m_{pq}^{\nabla y}$  have to be evaluated to arrive at the interaction matrix. This in turn would require the computation of the image gradient terms  $I_x$  and  $I_y$ , an image processing step performed using derivative filters, which might introduce an imprecision in the computed values. In the following, it is shown that a clever application of the Green's theorem can help subvert the image gradients computation.

The Green's theorem helps to compute the integral of a function defined over a subdomain  $\pi$  of  $\mathbb{R}^2$  by transforming it into a line (curve/contour) integral over the boundary of  $\pi$ , denoted here as  $\partial\pi$ :

$$\iint_{\pi} \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_{\partial\pi} P dx + \oint_{\partial\pi} Q dy \quad (16)$$

With suitable choices of functions  $P$  and  $Q$ , we aim to transform the terms  $m_{pq}^{\nabla x}$  and  $m_{pq}^{\nabla y}$ . To compute  $m_{pq}^{\nabla x}$ , we let  $Q = x^p y^q w(\mathbf{x}) I(\mathbf{x})$  and  $P = 0$ . We have  $\frac{\partial P}{\partial y} = 0$  and

$$\frac{\partial Q}{\partial x} = p x^{p-1} y^q w(\mathbf{x}) I(\mathbf{x}) + x^p y^q \frac{\partial w}{\partial x} I(\mathbf{x}) + x^p y^q w(\mathbf{x}) I_x \quad (17)$$

Substituting this back into (16), we can write

$$\begin{aligned} \iint_{\pi} \left[ p x^{p-1} y^q w(\mathbf{x}) I(\mathbf{x}) + x^p y^q \frac{\partial w}{\partial x} I(\mathbf{x}) \right. \\ \left. + x^p y^q w(\mathbf{x}) I_x \right] dx dy = \oint_{\partial\pi} x^p y^q w(\mathbf{x}) I(\mathbf{x}) dy \end{aligned} \quad (18)$$

Recalling our compact notation in (14a) and rearranging (18), we obtain

$$\begin{aligned} m_{pq}^{\nabla x} &= - \iint_{\pi} \left( p x^{p-1} y^q w(\mathbf{x}) I(\mathbf{x}) + x^p y^q \frac{\partial w}{\partial x} I(\mathbf{x}) \right) dx dy \\ &\quad + \oint_{\partial\pi} x^p y^q w(\mathbf{x}) I(\mathbf{x}) dy \end{aligned}$$

Applying (1) to the first term in the RHS, we have

$$\begin{aligned} m_{pq}^{\nabla x} &= -p m_{p-1,q} - \iint_{\pi} x^p y^q \frac{\partial w}{\partial x} I(\mathbf{x}) dx dy \\ &\quad + \oint_{\partial\pi} x^p y^q w(\mathbf{x}) I(\mathbf{x}) dy \end{aligned} \quad (19)$$

In the same manner, the computation of the term  $m_{pq}^{\nabla y}$  is again simplified by employing the Green's theorem with  $P = x^p y^q w(\mathbf{x}) I(\mathbf{x})$  and  $Q = 0$ .

$$m_{pq}^{\nabla y} = -q m_{p,q-1} - \iint_{\pi} x^p y^q \frac{\partial w(x,y)}{\partial y} I(\mathbf{x}) dx dy - \oint_{\partial\pi} x^p y^q w(\mathbf{x}) I(\mathbf{x}) dx \quad (20)$$

The results (19) and (20) are generic, meaning there are no explicit conditions on the weighting except that the function is differentiable. Clearly, depending on the nature of the weighting chosen for the measured intensities in (1), different analytical results can be obtained. In the following, two variants of the interaction matrix are developed corresponding to two different choices for the spatial weighting.

#### A. Uniformly Weighted Photometric Moments (UWPM)

First, the interaction matrix is established by attributing the same importance to all the measured intensities on the image plane. These moments are obtained by simply fixing  $w(\mathbf{x}, t) = 1, \forall \mathbf{x} \in \pi$  leading to  $\frac{\partial w}{\partial x} = \frac{\partial w}{\partial y} = 0$ . Subsequently, (19) and (20) get reduced to

$$\begin{cases} m_{pq}^{\nabla x} = -p m_{p-1,q} + \oint_{\partial\pi} x^p y^q I(x,y) dy \\ m_{pq}^{\nabla y} = -q m_{p,q-1} - \oint_{\partial\pi} x^p y^q I(x,y) dx \end{cases} \quad (21)$$

The second terms in  $m_{pq}^{\nabla x}$  and  $m_{pq}^{\nabla y}$  are contour integrals along  $\partial\pi$ . These terms represent the contribution of information that enter and leave the image due to camera motion. They could be evaluated directly but for obtaining simple closed-form expressions, the conditions under which they vanish are studied. Let us denote  $I_{\partial\pi} = \oint_{\partial\pi} x^p y^q I(x,y) dy$ .

The limits  $y = y_m$  and  $y = y_M$  are introduced at the top and bottom of the image respectively (see Fig 1a). Since  $y(=y_M)$  is constant along  $C1$  and  $y(=y_m)$  is constant along  $C3$ , it is sufficient to integrate along  $C2$  and  $C4$ . Along  $C2$ ,  $y$  varies from  $y_M$  to  $y_m$  while  $x$  remains constant at  $x_M$ . Along  $C4$ ,  $y$  varies from  $y_m$  to  $y_M$  while  $x$  remains constant at  $x_m$ . Introducing these limits, we get

$$I_{\partial\pi} = x_M^p \int_{y_M}^{y_m} y^q I(x_M, y) dy + x_m^p \int_{y_m}^{y_M} y^q I(x_m, y) dy$$

If  $I(x_M, y) = I(x_m, y) = I, \forall y$ , then we have

$$I_{\partial\pi} = (x_M^p - x_m^p) I \int_{y_M}^{y_m} y^q dy$$

Since we want  $I_{\partial\pi} = 0$ , the only solution is to have  $I = 0$ , that is when the acquired image is surrounded by a uniformly colored  $black^2$  background. This assumption, named information persistence (IP) was already implicitly done in

<sup>2</sup>or white with the intensity  $I$  redefined to  $I_{\max} - I$  and the rest of the developments remain identical.

[17] [24]. It does not need not be strictly enforced. In fact, mild violations of the IP assumption were deliberately introduced in experiments (refer IV-B) and this was quite acceptable in most cases, as evidenced by our results. This assumption gets naturally eliminated when appropriate weighting functions are introduced in the moments formulation as shown in II-B.

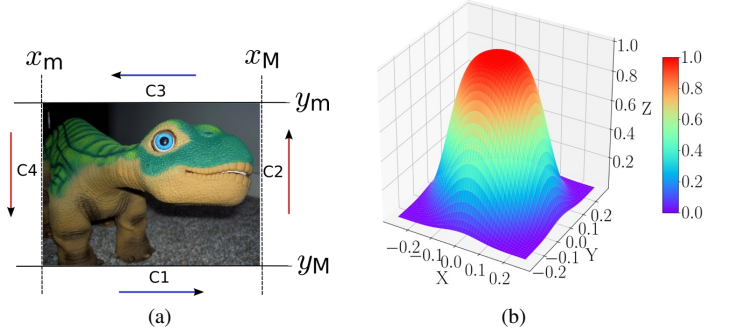


Fig. 1. a) Evaluation of contour integrals in the interaction matrix developments, b) Custom exponential function  $w(x,y) = \exp^{-650(x^2+y^2)^2}$  in the domain  $-0.4 \leq x \leq 0.4$  and  $-0.3 \leq y \leq 0.3$ . Gradual reduction in importance from maximum (dark red) in the centre outwards to minimum (blue) at the edges

With the same line of reasoning, the contour integral in  $m_{pq}^{\nabla y}$  also vanishes. Then (21) transforms to the following simple form:

$$\begin{cases} m_{pq}^{\nabla x} = -p m_{p-1,q} \\ m_{pq}^{\nabla y} = -q m_{p,q-1} \end{cases} \quad (22)$$

Substituting (22) into (15), we get the final closed form expression for the interaction matrix.

$$\begin{aligned} L_{m_{pq}}^{v_x} &= -A(p+1)m_{pq} - Bpm_{p-1,q+1} - Cpm_{p-1,q} \\ L_{m_{pq}}^{v_y} &= -Aqm_{p+1,q-1} - B(q+1)m_{p,q} - Cqm_{p,q-1} \\ L_{m_{pq}}^{v_z} &= A(p+q+3)m_{p+1,q} \\ &\quad + B(p+q+3)m_{p,q+1} + C(p+q+2)m_{p,q} \\ L_{m_{pq}}^{\omega_x} &= qm_{p,q-1} + (p+q+3)m_{p,q+1} \\ L_{m_{pq}}^{\omega_y} &= -pm_{p-1,q} - (p+q+3)m_{p+1,q} \\ L_{m_{pq}}^{\omega_z} &= pm_{p-1,q+1} - qm_{p+1,q-1} \end{aligned} \quad (23)$$

The interaction matrix in (23) has a form which is exactly identical to those developed earlier for the geometric moments [6]. A consistency with previously developed results is thus observed even though the method used for the modelling developments differ completely from [6]. Consequently, all the useful results available in the state of the art with regards to the developments of visual features [7] [8] are applicable as they are for the proposed photometric moments. Unlike (15), the image gradients do not appear anymore in the interaction matrix. Their computation is no longer necessary. The developments presented have led to the elimination of this image processing step required by pure luminance-based visual servoing [14]. The computation of the interaction matrix is now reduced to a simple and straight-forward computation of the moments on the image plane. Note also that in order to calculate  $L_{m_{pq}}$ , only moments of order upto  $p+q+1$

are required. In addition, we note that as usual in IBVS, the interaction matrix components corresponding to the rotational degrees of freedom are free from 3D parameters.

### B. Weighted Photometric Moments (WPM)

In order to remove the IP assumption we do not attribute anymore an equal contribution to all the measured intensities ( $w(\mathbf{x}) \neq 1, \forall \mathbf{x} \in \partial\pi$ ), as was done in Sec II-A. Instead, a lesser importance is attributed to peripheral pixels, on which the appearance and disappearance effects are pronounced. To achieve this, the spatial weighting function is made to attribute maximal importance to the pixels in the area around the image center and smoothly reducing it radially outwards towards 0 at the image periphery. If  $w(x, y) = 0, \forall \mathbf{x} \in \partial\pi$ , this still ensures  $I_{\partial\pi} = 0$  obviating the need to have any explicit IP assumption anymore.

*Weighting scheme:* The standard logistic function  $l(x) = \frac{1}{1+e^{-x}}$  smoothly varies between 0 and 1 and has simple derivatives. It is a standard function that is used in machine learning. However, if used to design  $w(\mathbf{x})$ , it is straight-forward to check that the interaction matrix cannot be expressed as functions of the weighted photometric moments. To achieve this, we propose to use functions with the general structure:

$$\mathcal{F}(x) = K \exp^{-p(x)} \quad (24)$$

with  $p(x) = a_0 + a_1x + \frac{1}{2}a_2x^2 + \frac{1}{3}a_3x^3 + \dots + \frac{1}{n}a_nx^n$ . Indeed, functions of this structure possess the interesting property that their derivatives can be expressed in terms of the function itself. It is given by:

$$\mathcal{F}'(x) = -K \exp^{-p(x)} p'(x) = -p'(x)\mathcal{F}(x)$$

with  $p'(x) = a_1 + a_2x + a_3x^2 + \dots + a_nx^{n-1}$ . In line with the above arguments, we propose the following custom exponential function (see Fig 1b)

$$w(x, y) = K \exp^{-a(x^2+y^2)^2} \quad (25)$$

where  $K$  is the maximum value that  $w$  can attain and  $a$  can be used to vary the area which receives maximal and minimal weights respectively. This choice allows the interaction matrix to be obtained directly in closed-form as a function of the weighted moments. Therefore, no additional computational overheads are introduced since nothing other than weighted moments upto a specific order are required. In addition, the symmetric function to which the exponential is raised ensures that the spatial weighting does not alter the behaviour of weighted photometric moments to planar rotations.

The spatial derivatives of (25) are as follows:

$$\begin{cases} \frac{\partial w}{\partial x} = -4ax(x^2 + y^2) w(\mathbf{x}) \\ \frac{\partial w}{\partial y} = -4ay(x^2 + y^2) w(\mathbf{x}) \end{cases} \quad (26)$$

Substituting (26) into (19) and (20), we obtain

$$\begin{cases} m_{pq}^{\nabla x} = -p m_{p-1,q} + 4a(m_{p+3,q} + m_{p+1,q+2}) \\ m_{pq}^{\nabla y} = -q m_{p,q-1} + 4a(m_{p,q+3} + m_{p+2,q+1}) \end{cases} \quad (27)$$

By combining (27) with the generic form in (15), the interaction matrix of photometric moments  ${}^w\mathbf{L}_{m_{pq}}$  weighted with the radial function (25) is obtained.

$${}^w\mathbf{L}_{m_{pq}} = \begin{bmatrix} {}^wL_{m_{pq}}^{v_x} & {}^wL_{m_{pq}}^{v_y} & {}^wL_{m_{pq}}^{v_z} & {}^wL_{m_{pq}}^{\omega_x} & {}^wL_{m_{pq}}^{\omega_y} & {}^wL_{m_{pq}}^{\omega_z} \end{bmatrix} \quad (28)$$

with

$$\begin{aligned} {}^wL_{m_{pq}}^{v_x} &= L_{m_{pq}}^{v_x} + 4aA(m_{p+4,q} + m_{p+2,q+2}) \\ &\quad + 4aB(m_{p+3,q+1} + m_{p+1,q+3}) \\ &\quad + 4aC(m_{p+3,q} + m_{p+1,q+2}) \\ {}^wL_{m_{pq}}^{v_y} &= L_{m_{pq}}^{v_y} + 4aA(m_{p+3,q+1} + m_{p+1,q+3}) \\ &\quad + 4aB(m_{p,q+4} + m_{p+2,q+2}) \\ &\quad + 4aC(m_{p,q+3} + m_{p+2,q+1}) \\ {}^wL_{m_{pq}}^{v_z} &= L_{m_{pq}}^{v_z} - 4aA(m_{p+5,q} + 2m_{p+3,q+2} + m_{p+1,q+4}) \\ &\quad - 4aB(m_{p+4,q+1} + 2m_{p+2,q+3} + m_{p,q+5}) \\ &\quad - 4aC(m_{p+4,q} + 2m_{p+2,q+2} + m_{p,q+4}) \\ {}^wL_{m_{pq}}^{\omega_x} &= L_{m_{pq}}^{\omega_x} - 4a(m_{p+4,q+1} + 2m_{p+2,q+3} + m_{p,q+3} \\ &\quad + m_{p+2,q+1} + m_{p,q+5}) \\ {}^wL_{m_{pq}}^{\omega_y} &= L_{m_{pq}}^{\omega_y} + 4a(m_{p+3,q} + m_{p+1,q+2} + m_{p+5,q} \\ &\quad + 2m_{p+3,q+2} + m_{p+1,q+4}) \\ {}^wL_{m_{pq}}^{\omega_z} &= L_{m_{pq}}^{\omega_z} = pm_{p-1,q+1} - qm_{p+1,q-1} \end{aligned}$$

We note that the interaction matrix can be expressed as a matrix sum

$${}^w\mathbf{L}_{m_{pq}} = \mathbf{L}_{m_{pq}} + 4a\mathbf{L}_w \quad (29)$$

where  $\mathbf{L}_{m_{pq}}$  has the same form as (23). We note however that the moments are now computed using the weighting function in (25). The matrix  $\mathbf{L}_w$  is tied directly to the weighting function. Of course if  $a = 0$  which means  $w(\mathbf{x}) = 1, \forall \mathbf{x} \in \pi$ , we find  ${}^w\mathbf{L}_{m_{pq}} = \mathbf{L}_{m_{pq}}$ .

To compute  $\mathbf{L}_{m_{pq}}$ , moments of order upto  $(p+q+1)$  are required whereas  $\mathbf{L}_w$  is a function of moments  $m_{tu}$ , where  $t+u \leq p+q+5$ . This is in fact a resultant of the term  $(x^2+y^2)^2$  to which the exponential is raised (see (25)).

On observation of the last component of  ${}^w\mathbf{L}_{m_{pq}}$ , we see that it does not contain any new terms when compared to (23). That is, the weighting function has not induced any extra terms, thus retaining the invariance of the classical moment invariants to optic axis rotations. This outcome was of course desired from the symmetry of the weighting function. On the other hand, if we consider the other five components, additional terms are contributed by the weighting function. As a result, moment polynomials developed from the classical moments [7] will not be invariant to translational motions when used with WPM. Thus, there is a need to develop new invariants for use with WPM such that they would retain their invariance to translations. This is an open problem that is not dealt with in this paper. Finally and as usual, the components of the interaction matrix corresponding to the rotational motions are still free from any 3D parameters.

Weighted photometric moments allow visual servoing on scenes prone to appearance and disappearance effects. Moreover, the interaction matrix has been developed in closed-form in order to facilitate detailed stability and robustness analyses.

The above developments would be near-identical for other weighting function choices of the form given by (24) [25].

### III. VISUAL FEATURES AND CONTROL SCHEME

The photometric moments are image-based measurements  $\mathbf{m}(t) = (m_{00}(t), m_{10}(t), m_{01}(t), \dots)$  obtained from the image  $\mathbf{I}(t)$ . To control  $n$  ( $\leq 6$ ) degrees of freedom of the robot, a large set of  $k$  ( $> n$ ) individual photometric moments could be used as input  $\mathbf{s}$  to the control scheme:  $\mathbf{s} = \mathbf{m}(t)$ . However, this would lead to redundant features, for which it is well known that, at best, only the local asymptotic stability can be demonstrated [3]. That is why we prefer to use the same strategy as in [6]–[8], that is, from the set of available measurements  $\mathbf{m}(t)$ , we design a set of  $n$  visual features  $\mathbf{s} = \mathbf{s}(\mathbf{m}(t))$  so that  $\mathbf{L}_s$  is of full rank  $n$  and has nice decoupling properties. The interaction matrix  $\mathbf{L}_s$  can easily be obtained from the matrices  $\mathbf{L}_{m_{pq}} \in \mathbb{R}^{1 \times 6}$  modelled in the previous section. Indeed, we have:

$$\mathbf{L}_s = \frac{\partial \mathbf{s}}{\partial \mathbf{m}} \mathbf{L}_m \quad (30)$$

where  $\mathbf{L}_m$  is the matrix obtained by stacking the matrices  $\mathbf{L}_{m_{pq}}$ . Then, the control scheme with the most basic and classical form has been selected [3]:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L}}_s^{-1} (\mathbf{s} - \mathbf{s}^*) \quad (31)$$

where  $\mathbf{s}^* = \mathbf{s}(\mathbf{m}^*)$  and  $\widehat{\mathbf{L}}_s$  is an estimation or an approximation of  $\mathbf{L}_s$ . Such an approximation or estimation is indeed necessary since, as detailed in the previous section, the translational components of  $\mathbf{L}_{m_{pq}}$  are function of the 3D parameters  $A_{pq}$  describing the depth map of the scene. Classical choices are  $\widehat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}(t), \widehat{\mathbf{Z}}(t))$  where  $\mathbf{Z} = (A, B, C)$  when an estimation of  $\mathbf{Z}$  is available,  $\widehat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}(t), \widehat{\mathbf{Z}}^*)$ , or even  $\widehat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}^*, \widehat{\mathbf{Z}}^*)$ . Another classical choice is to use the mean  $\widehat{\mathbf{L}}_s = \frac{1}{2} (\mathbf{L}_s(\mathbf{s}(t), \widehat{\mathbf{Z}}(t)) + \mathbf{L}_s(\mathbf{s}^*, \widehat{\mathbf{Z}}^*))$  or  $\widehat{\mathbf{L}}_s = \frac{1}{2} (\mathbf{L}_s(\mathbf{s}(t), \widehat{\mathbf{Z}}^*) + \mathbf{L}_s(\mathbf{s}^*, \widehat{\mathbf{Z}}^*))$  since it was shown to be efficient for very large camera displacements [26].

With such a control scheme, it is well known that the global asymptotic stability (GAS) of the system in the Lyapunov sense is ensured if the following sufficient condition holds [3]:

$$\mathbf{L}_s \widehat{\mathbf{L}}_s^{-1} > 0 \quad (32)$$

Of course, in case  $\widehat{\mathbf{L}}_s = \mathbf{L}_s$ , the system is GAS if  $\mathbf{L}_s$  is never singular, and a perfect decoupled exponential decrease of the error  $\mathbf{s} - \mathbf{s}^*$  is obtained. Such a perfect behavior is not obtained as long as  $\widehat{\mathbf{L}}_s \neq \mathbf{L}_s$ , but the error norm will decrease and the system will converge if condition (32) is ensured. This explains the fact that a non planar scene can be considered in practice (see Section V-D), even if the modelling developed in the previous section was limited to the planar case.

#### A. Control of SCARA motions

Photometric moments-based visual features can be used to control not only the subset of  $\mathbb{SE}(3)$  motions considered in [17] but also full 6 dof motions. In the former case, the robot is configured for SCARA (3T+1R,  $n = 4$ ) type actuation to

control only the planar translation, translation along the optic axis and rotation around the optic axis. The camera velocity is thus reduced to  $\mathbf{v}_{c_r} = (v_x, v_y, v_z, \omega_z)$ . Similarly to [7], the following set of 4 visual features is used to control these 4 dofs.

$$\mathbf{s}_r = (x_n, y_n, a_n, \alpha) \quad (33)$$

where  $x_n = x_g a_n$ ,  $y_n = y_g a_n$ ,  $a_n = Z^* \sqrt{\frac{m_{00}^*}{m_{00}}}$  with  $x_g = m_{10}/m_{00}$  and  $y_g = m_{01}/m_{00}$  the centre of gravity coordinates,  $Z^*$  the desired depth and finally  $\alpha = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right)$  is made of centred moments given by:

$$\begin{cases} \mu_{20} = m_{20} - m_{00} x_g^2 \\ \mu_{02} = m_{02} - m_{00} y_g^2 \\ \mu_{11} = m_{11} - m_{00} x_g y_g \end{cases} \quad (34)$$

From the simple relations between  $\mathbf{s}_r$  and  $m_{pq}$ , ( $p + q < 3$ ), it is quite simple to determine the analytical form of the interaction matrix  $\mathbf{L}_{s_r}$  using (30). When the target is parallel to the image plane ( $A = B = 0$ ), the following sparse matrix is obtained for UWPM.

$$\mathbf{L}_{s_r} = \begin{bmatrix} \mathbf{L}_{x_n} \\ \mathbf{L}_{y_n} \\ \mathbf{L}_{a_n} \\ \mathbf{L}_{\alpha} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & y_n \\ 0 & -1 & 0 & -x_n \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (35)$$

Let us note that the current value of the depth does not appear anywhere in  $\mathbf{L}_{s_r}$  and only the desired value  $Z^*$  intervenes indirectly through  $x_n$  and  $y_n$ , and thus in  $\mathbf{L}_{s_r}$ . This nice property and the sparsity in (35) justify the choice of  $\mathbf{s}_r$ . Following the line of analysis at the start of this section, we infer that the control law using  $\widehat{\mathbf{L}}_{s_r} = \mathbf{L}_{s_r}(\mathbf{s}_r(t), Z^*)$  is GAS since  $\mathbf{L}_{s_r}$  is always of full rank 4 and  $\mathbf{L}_{s_r} \widehat{\mathbf{L}}_{s_r}^{-1} = \mathbb{I}_4$  when  $\widehat{Z} = Z^*$ .

Let us now consider the more general case where  $\widehat{Z} \neq Z^*$ . From (35), it is straight-forward to obtain

$$\mathbf{L}_s \widehat{\mathbf{L}}_s^{-1} = \begin{bmatrix} 1 & 0 & 0 & Y \\ 0 & 1 & 0 & X \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (36)$$

where  $Y = (\frac{\widehat{Z}^*}{Z^*} - 1)y_n$  and  $X = (1 - \frac{\widehat{Z}^*}{Z^*})x_n$ . The eigen values of the symmetric part of the above matrix product are given by  $\lambda = \{1, 1, 1 \pm \frac{\sqrt{X^2 + Y^2}}{2}\}$ . For (32) to hold, all eigen values have to be positive, that is,  $\frac{\sqrt{X^2 + Y^2}}{2} < 1 \Leftrightarrow X^2 + Y^2 < 4$ . Back-substitution of  $X$  and  $Y$  yields the following bounds for system stability:

$$1 - \frac{2}{\sqrt{x_n^2 + y_n^2}} < \frac{\widehat{Z}^*}{Z^*} < 1 + \frac{2}{\sqrt{x_n^2 + y_n^2}} \quad (37)$$

which are easily ensured in practice since  $x_n$  and  $y_n$  are small (0.01 typically).

Let us now consider the case where  $\widehat{\mathbf{L}}_s = -\mathbb{I}_4$ , which is a coarse approximation. In that case, we obtain

$$\mathbf{L}_s \widehat{\mathbf{L}}_s^{-1} = \begin{bmatrix} 1 & 0 & 0 & -y_n \\ 0 & 1 & 0 & x_n \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (38)$$

Then, proceeding as previously leads to the following condition for GAS

$$x_n^2 + y_n^2 < 4 \quad (39)$$

which, once again, is always ensured in practice. Note that these satisfactory theoretical results have not been reported previously and are an original contribution of this work. Unfortunately, exhibiting similar conditions for the WPM case is not so easy since the first three columns of  $\mathbf{L}_{s_r}$  are not as simple as (35) due to the loss of invariance property of WPM.

### B. 6 dof control

To control all the 6 dof, two more features in addition to (33) are required. In moments-based VS methods, these features are chosen as ratios of moment polynomials which are invariant to 2D translations, planar rotation and scale. In [7], [8], several moment invariants-based visual features have been introduced. In principle, all these previous results could be adopted for use with the photometric moments proposed in this work. Certainly, an exhaustive exploration of all these choices is impractical. Based on several simulations and experimental convergence trials (see [25]), the following visual feature introduced in [8] was selected:

$$r = \Phi_1 / \Phi_2 \quad (40)$$

$$\text{with } \begin{cases} \Phi_1 = 3\tilde{\mu}_{30}\tilde{\mu}_{12} + \tilde{\mu}_{30}^2 + 3\tilde{\mu}_{03}\tilde{\mu}_{21} + \tilde{\mu}_{03}^2 \\ \Phi_2 = -\tilde{\mu}_{30}\tilde{\mu}_{12} + \tilde{\mu}_{21}^2 - \tilde{\mu}_{03}\tilde{\mu}_{21} + \tilde{\mu}_{12}^2 \end{cases} \quad (41)$$

where  $\tilde{\mu}_{pq}$  is the shifted moment of order  $p + q$  with respect to shift point  $\mathbf{x}_{sh}(x_{sh}, y_{sh})$  defined by [8]:

$$\tilde{\mu}_{pq} = \iint (x - x_g + x_{sh})^p (y - y_g + y_{sh})^q w(\mathbf{x}) I(\mathbf{x}) \, dx \, dy \quad (42)$$

where the shift point coordinates  $(x_{sh}, y_{sh})$  are defined by

$$\begin{cases} x_{sh} = \sqrt{m_{00}} \cos \theta \\ y_{sh} = \sqrt{m_{00}} \sin \theta \end{cases} \quad (43)$$

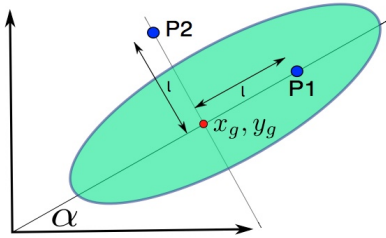


Fig. 2. Shift points  $\mathbf{P}_1(\mathbf{x}_g + \mathbf{x}_{sh1})$  and  $\mathbf{P}_2(\mathbf{x}_g + \mathbf{x}_{sh2})$  with respect to which the shifted moments are computed.

As shown in Fig 2, one shift point is selected along the major orientation ( $\theta = \alpha$ ) and the second point orthogonal to the previous ( $\theta = \alpha + \frac{\pi}{2}$ ) such that we have :  $\mathbf{P}_1[x_g + \sqrt{m_{00}} \cos(\alpha), y_g + \sqrt{m_{00}} \sin(\alpha)]$  and  $\mathbf{P}_2[x_g + \sqrt{m_{00}} \cos(\alpha + \frac{\pi}{2}), y_g + \sqrt{m_{00}} \sin(\alpha + \frac{\pi}{2})]$ . To sum up, the shifted moments in (42) are computed with respect to  $\mathbf{P}_1$  and  $\mathbf{P}_2$ , resulting in two different sets of shifted moments. Then, the feature in (40) is computed employing these two sets of moments to derive two corresponding visual features  $r_{P_1}$  and  $r_{P_2}$ . Therefore, the

following set of visual features for controlling the 6 dof is obtained:

$$\mathbf{s} = (x_n, y_n, a_n, r_{P_1}, r_{P_2}, \alpha) \quad (44)$$

The interaction matrix developments of  $r_{P_1}$  and  $r_{P_2}$  are provided in Appendix A. When UWPM are used, the interaction matrix  $\mathbf{L}_s$  exhibits the following sparse structure when the sensor and target planes are parallel.

$$\mathbf{L}_s^{\parallel} = \begin{bmatrix} \mathbb{I}_3 & D \\ \mathbf{0}_3 & E \end{bmatrix} \quad (45)$$

$$\text{with } D = \begin{bmatrix} L_{x_n}^{\omega_x} & L_{x_n}^{\omega_y} & y_n \\ L_{y_n}^{\omega_x} & L_{y_n}^{\omega_y} & -x_n \\ L_{a_n}^{\omega_x} & L_{a_n}^{\omega_y} & 0 \end{bmatrix} \text{ and } E = \begin{bmatrix} L_{r_{P_1}}^{\omega_x} & L_{r_{P_1}}^{\omega_y} & 0 \\ L_{r_{P_2}}^{\omega_x} & L_{r_{P_2}}^{\omega_y} & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The matrix  $E$  is non-singular if its left  $2 \times 2$  submatrix has a non-zero determinant. When the interaction matrix is computed with moments from shift points ( $\mathbf{P}_1 \neq \mathbf{P}_2$ ) as described above, this condition is effortlessly ensured. As a result, the interaction matrix  $\mathbf{L}_s^{\parallel}$  is non-singular [8]. On the other hand, when the features are built from WPM, the sparsity in (45) cannot be achieved anymore. This is because  $\mathbf{L}_s$  has a more complex form, except for its last column which remains exactly the same (since behaviour with respect to optic axis rotations is not altered). Nevertheless, the obtained results were quite satisfactory for a variety of scenes and camera displacements, as shown in the next section.

## IV. VALIDATION RESULTS FOR UWPM

### A. Modelling Validation and Comparison to Pure Luminance

In this section, simulation results of 6 dof positioning tasks are presented to demonstrate the correctness of the modelling of the UWPM proposed in Sec.II-A and to compare their behavior to pure luminance. The initial and desired images are shown in Figs.3a and 3b respectively. The background is empty without appearance or disappearance of scene portions in the camera view. The initial pose is chosen far away from the desired one such that the image overlap is small. The displacements required for convergence are a translation of  $\mathbf{t} = [1.0m, 1.0m, 1.0m]$  and a rotation of  $\mathbf{R} = [25^\circ, 10^\circ, 55^\circ]$ .

The control law in (31) is used with  $\widehat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}(t), \mathbf{Z}(t))$ . This control law is expected to result in a pure exponential decrease of the errors to 0. In simulation, the depths  $\mathbf{Z}(t)$  are readily available from ground truth and need not be estimated. A gain of  $\lambda = 1.0$  was used for this experiment.

As seen from Fig 3c, a perfect exponential decrease of the errors is indeed obtained as expected. Furthermore, the camera traces a straight-forward path to the goal pose as shown in Fig 3d. This demonstrates the validity of the modelling steps and the design of the visual features. Let us note that no image processing (image matching or visual tracking) were used with the photometric moments in the reported experiments.

*Comparison to pure luminance:* Then, the same control law configuration was tested using pure luminance directly as visual feature, that is using  $\mathbf{v}_c = -\lambda \widehat{\mathbf{L}}_I^+ (\mathbf{I} - \mathbf{I}^*)$ . The velocity profiles generated are shown in Figure 4c. The pure luminance experiment is not successful as it results in an enormous final error  $\mathbf{e} \approx 10^9$ , as seen from Fig 4a. In direct



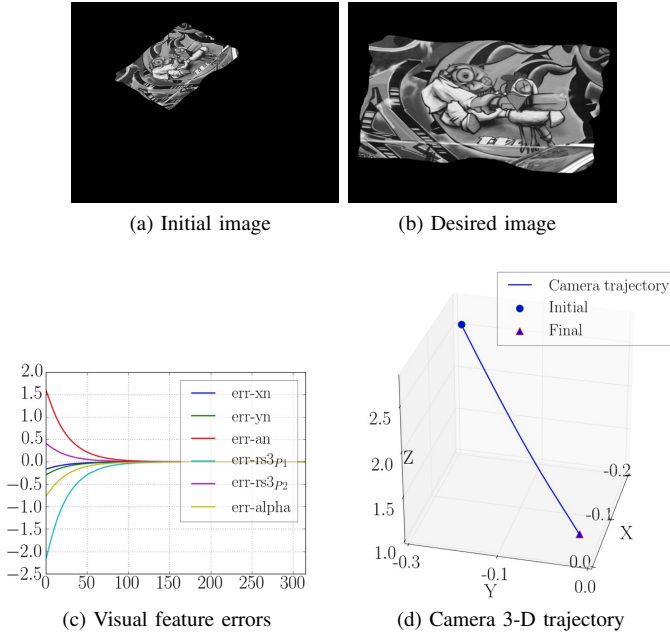


Fig. 3. Simulation results with UWPM in perfect conditions

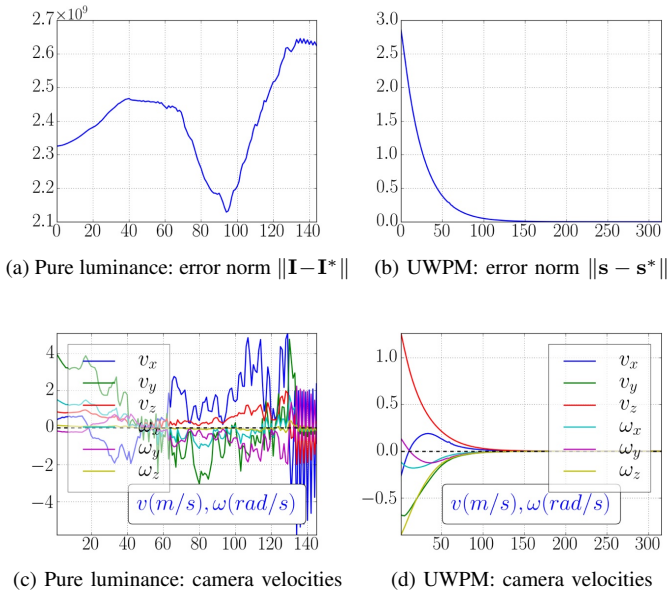


Fig. 4. Comparison of UWPM and pure luminance VS

contrast, the control velocities computed using the photometric moments (see Fig 4d) ensures a clear monotonous decrease of the feature error norm, as shown in Fig 4b. This evidence supports our assertion that the visual features based on the photometric moments have a larger convergence domain than the pure luminance features. With photometric moments, we can appreciate the direct correspondence between each feature and the degree of freedom it intends to control, a property highly desired in visual servo control. The camera view and evolution of the error images in the above cases are shown in the video supplement.

## B. Experimental Results with UWPM

Experiments were performed at video rate on a Viper850 6 dof robot. Unlike in Sec IV-A, mild violations of the IP assumption are deliberately allowed. The photometric moments are tested first on SCARA-type motions and then with 6 dof.

1) *SCARA motions*: For this experiment, the features in (33) are used with their current interaction matrix  $\hat{\mathbf{L}}_s = \hat{\mathbf{L}}_s(\mathbf{s}(t), \hat{\mathbf{Z}}^*)$ , with  $\hat{\mathbf{Z}}^* = (0, 0, 1/\hat{Z}^*)$ ,  $\hat{Z}^*$  roughly approximated with depth value at the desired pose. A gain of  $\lambda = 1.5$  was used. The desired image is shown in Figure 5b. The initial pose is chosen such that the image in 5a is observed by the camera. The target is placed such that very small portions of its corners are slightly outside the field of view (see Fig 5a). Furthermore, the background is not perfectly black, thereby non-zero.

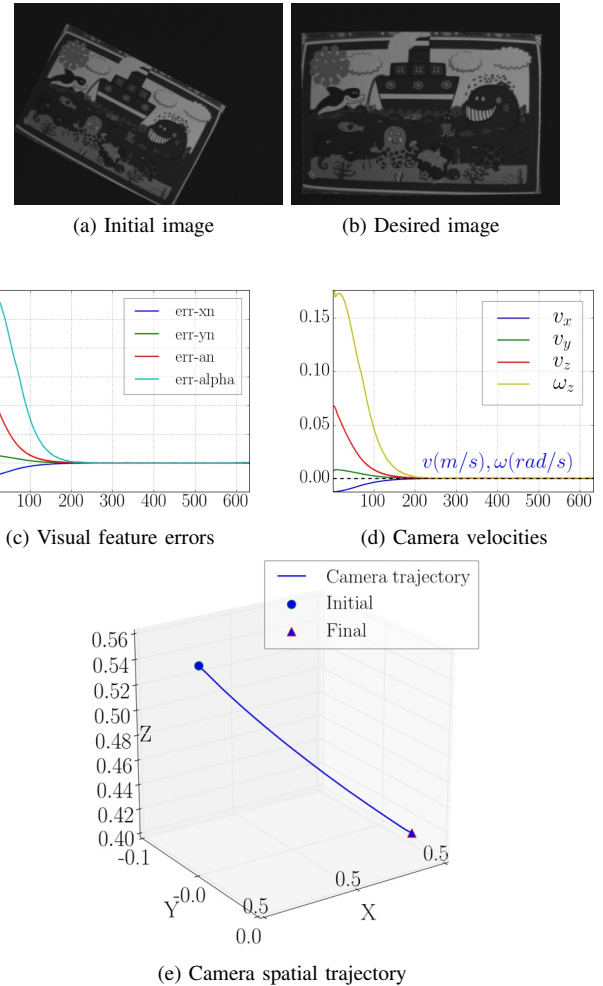


Fig. 5. Experimental results in SCARA mode actuation pertaining to Section IV-B1.

The generated control velocities are shown in Fig 5d. The decrease of the errors is purely exponential except for the small blip during the very few initial iterations, as shown in Fig 5c. This blip is due to the small portions of the target which enter the image and introduce a small disturbance in those initial iterations. The overall behavior is satisfactory with the camera

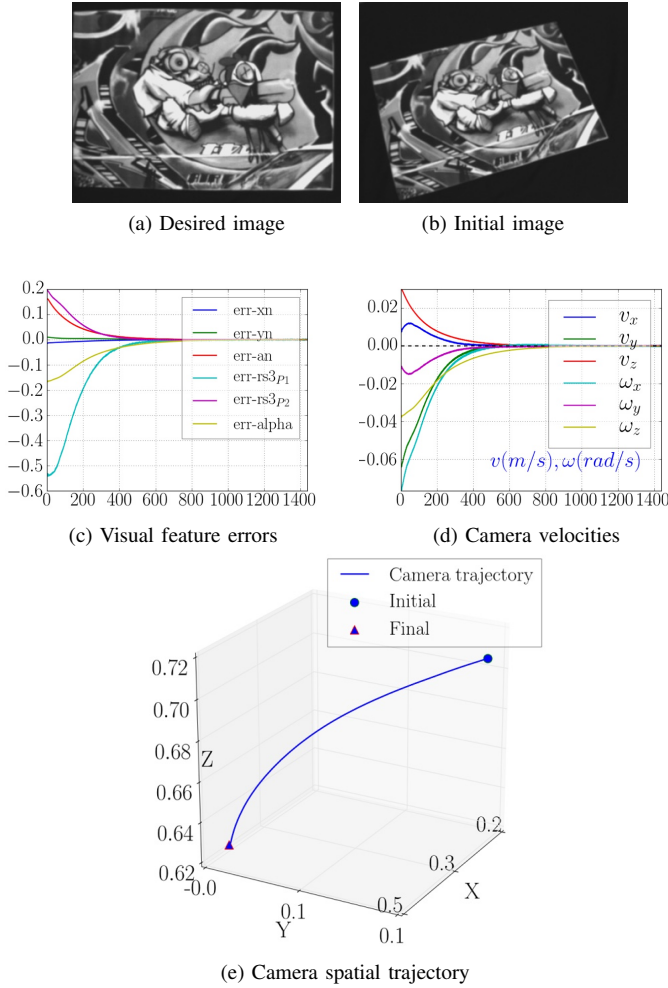


Fig. 6. 6 dof experimental results pertaining to Section IV-B2.

following a straight-forward trajectory to the goal as seen from Figure 5e.

2) *6 dof positioning tasks*: The desired pose of the camera is fronto-parallel, about 0.5m vertically above the target where the learnt image is as shown in Fig 6a. For the initial pose, the robot is steered far from the desired pose. At this initial pose, the sensor and target planes are not parallel anymore. The image acquired at this initial pose is shown in Fig 6b. The robot has to undergo a reasonably complex motion employing all its dof to achieve this task. The translational displacement is  ${}^c\mathbf{t}_{c*} = [26.0\text{cm}, 13.25\text{cm}, -8.00\text{cm}]$  and the rotational displacement is  ${}^c\mathbf{R}_{c*} = [7.0^\circ, -21.16^\circ, 17.10^\circ]$ . For the control law, the depth was measured once offline and approximated roughly at  $\hat{Z}^* = 0.5\text{m}$ . A constant gain of  $\lambda = 0.2$  was used. The set of visual features in (44) is used with the interaction matrix at the desired configuration  $\hat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}^*, \hat{\mathbf{Z}}^*)$  with  $\hat{\mathbf{Z}}^* = (0, 0, 1/\hat{Z}^*)$ .

It can be observed from Fig 6c that the decrease in errors is highly satisfactory while we recall that only the interaction matrix at the desired configuration and approximate depth were employed. The generated velocity profiles are also smooth as shown in Fig.6d. Clearly, the camera spatial trajectory is close to a geodesic as shown in Figure IV-B2. Further, an

accuracy of  $[-0.56\text{mm}, -0.08\text{mm}, 0.14\text{mm}]$  in translation and  $[-0.01^\circ, 0.04^\circ, -0.03^\circ]$  in rotation was obtained. The above experimental results showed results with UWPM where there are only mild violations of the IP assumption. Next, we show results on more general scenes with WPM where this restrictive assumption (black background) has been eliminated.

## V. VALIDATION RESULTS FOR WPM

For all the experiments presented in this section, the parameter  $K = 1$  is fixed, so maximum weight a pixel can have is 1. Then,  $a$  is chosen with a simple heuristic, that 40% of the image pixels will be assigned a weight greater than 0.5 and around 90% a weight greater than 0.01. This is straightforward to compute from the definition of  $w(x, y)$ . For an image resolution of  $640 \times 480$  for example, with  $K = 1$ ,  $a = 650$  satisfies the above simple heuristic. The surface of  $w(x, y)$  with these parameters is depicted in Fig. 1b. Let us note that the tuning of these parameters is not crucial. In our case, changing  $a$  by  $\pm 200$  does not introduce any drastic changes in the results.

### A. Validation of WPM

In this section, the modelling of WPM is validated using 6 dof positioning tasks in simulation. No specific backgrounds are considered anymore since the WPM designed in Section II-B are equipped to handle such scenarios. Comparisons to both the pure luminance feature and to moments without the weighting strategy are made.

The image learnt from the desired pose is shown in Fig 7b. In the image acquired from the initial robot pose (see Fig 7a), a large subset of pixels not present in the desired image have appeared. In fact, there is no clear distinction of which pixels constitute the background. These scenarios are more representative of camera-mounted robotic arms interacting with real world objects. For the control, the set of visual features (44) is adopted with the current interaction matrix  $\mathbf{L}_s(\mathbf{s}(t), \mathbf{Z}(t))$ . The depths are not estimated but available from the ground truth data. A gain of  $\lambda = 1.5$  was used for all the experiments. The resulting behaviour is very satisfactory. The errors in the visual features decrease exponentially as shown in Figures 7c and 7d. This confirms the correctness of the modelling steps used to obtain the interaction matrix of WPM. Naturally, the successful results also imply the correctness of the visual features obtained from the weighted moments.

*Comparison with UWPM*: For the comparison, the same experiment is repeated with the same control law but without the weighting strategy. In this case, the errors appear to decrease initially (see Figs 8a and 8b). However, after about 25 iterations the system diverges (see Fig 8c) and the servo is stopped after few iterations. As expected, the system in this case is clearly affected by the appearance and disappearance of parts of the scene.

*Comparison to pure luminance*: Next, we also compared the WPM with the pure luminance feature. Also in this case, the effect of the extraneous regions is severe and the control law does not converge to the desired pose. The generated velocities do not regulate the errors satisfactorily (see Fig 8d). The error

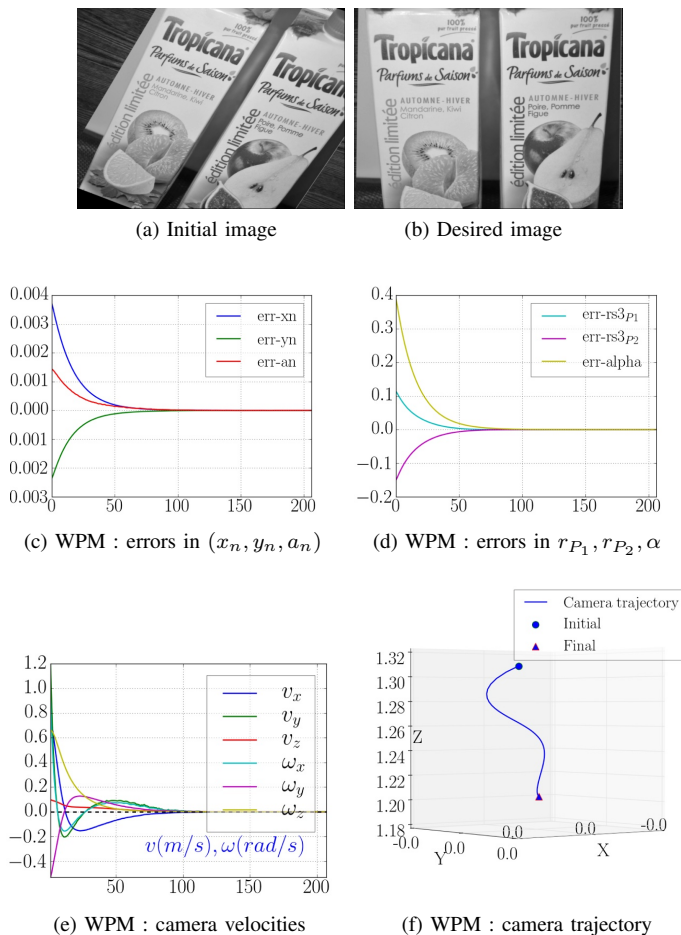


Fig. 7. Simulation V-A : 6 dof VS with WPM pertaining to Section V-A.

norm  $\|\mathbf{I} - \mathbf{I}^*\|$  starts to increase rapidly as shown in Fig 8f. This can be compared to the case of the WPM where the error norm decreases exponentially as shown in Figure 8e. Also, as mentioned previously, the visual features are redundant and there is no mapping of individual features to the actuated dof. The servoing behaviour depends on the profile of the cost function, which is dependent on all the intensities in the acquired image. The appearance and disappearance of scene portions thus also affects the direct visual servoing method. Thus, we see that the extraneous regions have resulted in the worst case effect namely non-convergence to the desired pose in both the UWPM as well as when using the pure luminance. Next, we discuss results obtained from servoing on a scene different from the one used in this experiment.

### B. Robustness to large rotations

In this simulation, we consider 4 dof and very large displacements such that large scene portions enter and leave the camera field of view (see Figures 9a and 9b).

A rotation of  $100^\circ$  around the optic axis and translational displacement of  $\mathbf{c}^* \mathbf{t}_c = [5cm, 4cm, 25cm]$  are required for convergence. For this experiment, the VS control law in (31) with the features in (33) is used with a gain of  $\lambda = 2$ . For this difficult task, the mean  $\widehat{\mathbf{L}}_s = \frac{1}{2} (\mathbf{L}_s(s(t), \widehat{\mathbf{Z}}^*) + \mathbf{L}_s(s^*, \widehat{\mathbf{Z}}^*))$

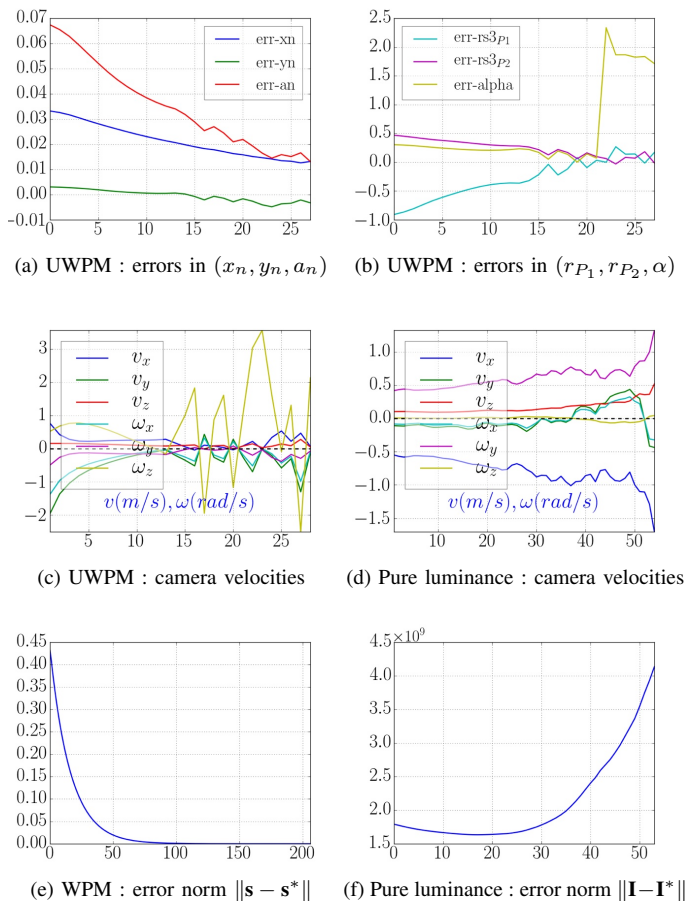
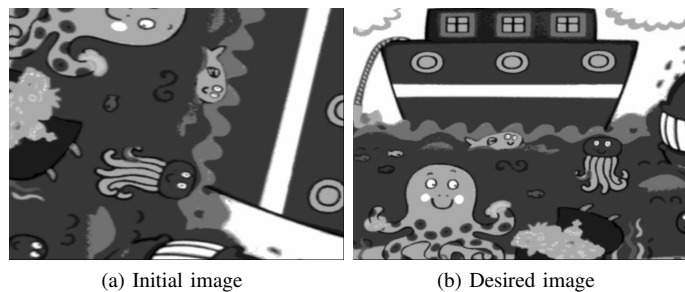


Fig. 8. Simulation V-A : 6 dof VS comparison to UWPM and pure luminance (see Fig. 7).

has been selected in the control scheme. Note that the depths are not updated at each iteration and only approximated using  $Z^* = 1$ . This choice was on purpose to show that online depth estimation is not necessary and an approximation of its value at the desired pose is sufficient for convergence. The visual servoing converged to the desired pose with an accuracy of  $0.29^\circ$  in rotation and  $[-0.07mm, -0.48mm, 0.61mm]$  in translation. The control velocities generated are shown in Fig.9d and the resulting Cartesian trajectories are shown in Fig.9e. This experiment demonstrates the robustness of the WPM to very large displacements even when there is appearance and disappearance of huge parts of the image. This affirms also that the convergence properties are improved with the proposed WPM.

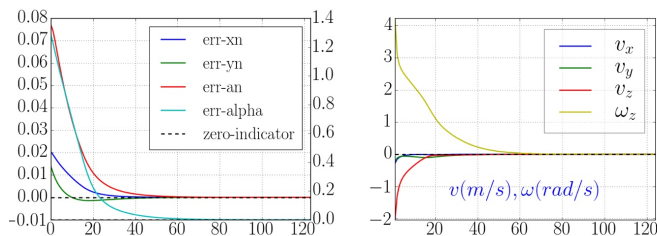
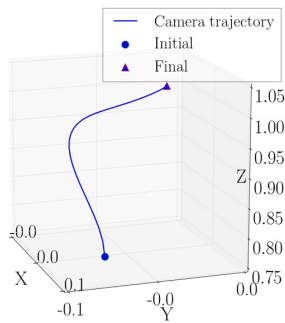
### C. Empirical Convergence Analysis

In this section, we compare through simulations the convergence domain of WPM with pure luminance and UWPM. For this, we considered the 4dof case as in [24]. Artificially generated synthetic scenes in which polygonal blocks are sprinkled at the image periphery were employed. As seen from Fig 10, this allows to simulate in varying degrees the appearance and disappearance of scene portions in the camera FOV. For this analysis, the desired pose to be attained is fixed



(a) Initial image

(b) Desired image

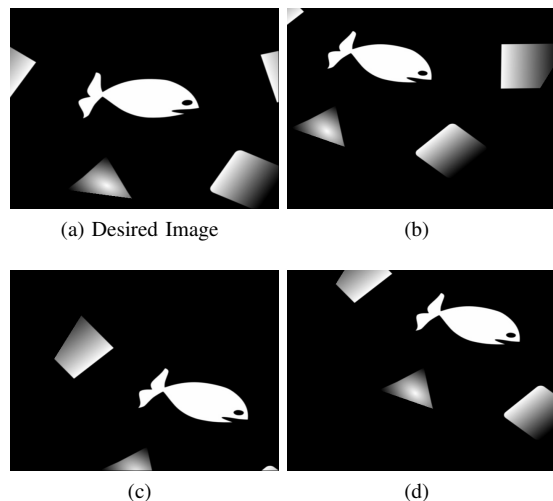
(c) Feature errors ( $(x_n, y_n, \alpha_n)$  on left  $y$ -axis,  $\alpha$  on right  $y$ -axis) (d) Camera velocities (left  $y$  axis for  $(v_x, v_y, v_z)$ , right  $y$  axis for  $\omega_z$ )

(e) Camera trajectory

Fig. 9. 4 dof simulation results under large rotations (see Section V-B).

at  $1.8m$ . Positioning tasks starting from 243 different initial poses consisting of 3 sets of 81 poses each, conducted at 3 different depths of  $1.8m$ ,  $1.9m$  and  $2.0m$  were considered. In all these initial poses, the camera is subjected to a rotation of  $25^\circ$  around the optic axis while the  $x$  and  $y$  translations vary from  $-0.2m$  to  $0.2m$ .

The interaction matrix  $\widehat{\mathbf{L}}_s = \mathbf{L}_s(\mathbf{s}^*, \widehat{\mathbf{Z}}^*)$  is chosen in the control scheme, just like in previous works on convergence analysis [14] [16]. We consider an experiment to have converged if the task error  $\|\mathbf{e}\|$  is reduced to less than  $1e^{-10}$  in a maximum of 300 iterations. In addition to this condition, we also impose that the SSD error defined by  $\mathbf{e}_{SSD} = \sum_{\mathbf{x}} [I(\mathbf{x}) - I^*(\mathbf{x})]^2 / N_{pix}$  between the final and learnt images is less than 1.0. This criterion ensures that a non-desired equilibrium point is not considered wrongly as converged. In the reported converged experiments, the final accuracy in pose is less than  $1mm$  for translations and less than  $1^\circ$  for the planar rotation. The UWPM met with failure in all the cases. No segmentation or thresholding is employed and the servo is subjected to appearance and disappearance effects at the image periphery. A dismal performance resulted as expected without the weighting strategy since the model is not equipped to handle the energy inflow and outflow at



(a) Desired Image

(b)

(c)

(d)

Fig. 10. Desired image in (a) and a sampling of different images from the 243 generated initial poses are shown in (b)-(d)

the image borders. The proposed WPM on the other hand achieved a convergence rate of 94.23%. The small number of failure cases were due to the system getting attracted to a global equilibrium different from the desired one. The pure luminance feature resulted in a convergence rate of 85.1%, still lower than the WPM. These good results are also due to the fact that the initial poses were not too far from the desired one. To demonstrate further the superiority of the WPM with

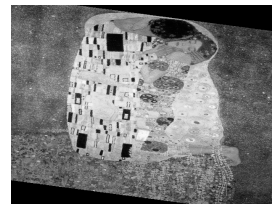


Fig. 11. Dense texture used in convergence analysis

respect to UWPM, the same set of experiments was repeated using a dense texture (see Fig.11), where the WPM yield a better result than non-weighted moments. The non-weighted moments have converged on an average only in 55% of the cases. Also note that this is different from the synthetic case at 0%, that is they were completely unable to handle the entry and exit of extraneous regions. In comparison, for WPM, only 3 cases failed to converge out of 243 total runs with a very satisfactory convergence rate of 98%. In fact, in the first two sets of experiments, WPM converged for all the generated poses yielding a 100% convergence rate. No convergence to any undesired equilibrium points were observed, thanks to the textured object. The final accuracies for all the converged experiments was less than  $1mm$  in translation and less than  $1^\circ$  in rotation. Based on the clear improvements in convergence rate, we conclude that WPM are effective as a solution to the problem of extraneous image regions and result in a larger convergence domain in comparison to classical non-weighted moments. We have finally to note that for larger lateral displacements, all methods fail since initial and desired

images do not share sufficient common information.

#### D. Robustness to non planar environments

In this section, visual servoing with WPM is demonstrated on a non planar scene with the Viper850 robot by considering 4 dof as previously. A realistic scenario is emulated by placing five 3D objects of varying shape, size and color in the scene as shown in Fig 12f. In the initial acquired image (see Fig 12a), 3 out of these 5 objects are not fully visible. The WPM were in fact conceived for use in such scenarios. Rotational displacement of  $-10^\circ$  around the optic axis and translations of  ${}^c\mathbf{t}_c = [1.5\text{cm}, 1\text{cm}, 8\text{cm}]$  are required for convergence. Once again, the mean interaction matrix  $\widehat{\mathbf{L}}_s = \frac{1}{2} \left( \mathbf{L}_s(\mathbf{s}(t), \widehat{\mathbf{Z}}^*) + \mathbf{L}_s(\mathbf{s}^*, \widehat{\mathbf{Z}}^*) \right)$  has been selected in the control scheme. The depth distributions in the scene are not estimated nor known apriori. An approximation  $\widehat{\mathbf{Z}}^* = (0, 0, 1/\widehat{Z}^*)$  with  $\widehat{Z}^* = 0.5\text{m}$  was used. A gain of  $\lambda = 0.4$  was employed. The control law generates camera velocities that decrease exponentially (see Fig 12d), which causes a satisfactory decrease in the feature errors (see Fig 12c). The average accuracy in positioning in translations is  $0.6\text{mm}$  while the rotational accuracy is  $0.15^\circ$ . The camera spatial trajectory is satisfactory as seen from Fig. 12e. The simplification (7) of planar scene introduced in the modelling (see Section II) is therefore a reasonable tradeoff of complexity, even if it is not possible to demonstrate that the sufficient stability condition (32) is ensured since  $\widehat{\mathbf{L}}_s \neq \mathbf{L}_s$ . This demonstrates the robustness of visual servoing with respect to (moderate) modelling approximations.

#### E. 6 dof experimental results

Several 6dof positioning experiments were conducted on the ViPER 850 robot. A representative one is presented below while the others can be consulted in [25]. For this experiment, the desired robot pose is such that the camera is at  $0.5\text{m}$  in a frontoparallel configuration in front of the target. The image learnt from this pose is shown in Fig 13b.

The initial pose is chosen such that the image in Fig 13a is observed. Let us note that *Lauren Bacal* present in the left part of the desired image is completely absent from the initial image. The corresponding difference image is shown in Fig 13c. There is no monotone segmented object and the assumption about uniform black background is clearly not valid in this case. Nominal displacements of  $[-0.35\text{cm}, -1.13\text{cm}, 6.67\text{cm}]$  in translation and  $[0.33^\circ, 1.05^\circ, 12.82^\circ]$  in rotation are required for convergence. The control law in (31) with the features in (44) is used, with  $\widehat{\mathbf{L}}_s$  as the mean of the desired and current interaction matrices. No pose estimation is performed and the depth is approximated roughly as  $0.5\text{m}$ . The appearance of new scene portions in the camera view from the left side of the image does not affect the convergence of the visual servo. This influx of information is handled gracefully thanks to the improved modelling used by the WPM. The error in features related to control of rotational motions is very satisfactory (see Fig13e). On the other hand, from the error decrease in features related to control of translational motions in Figure

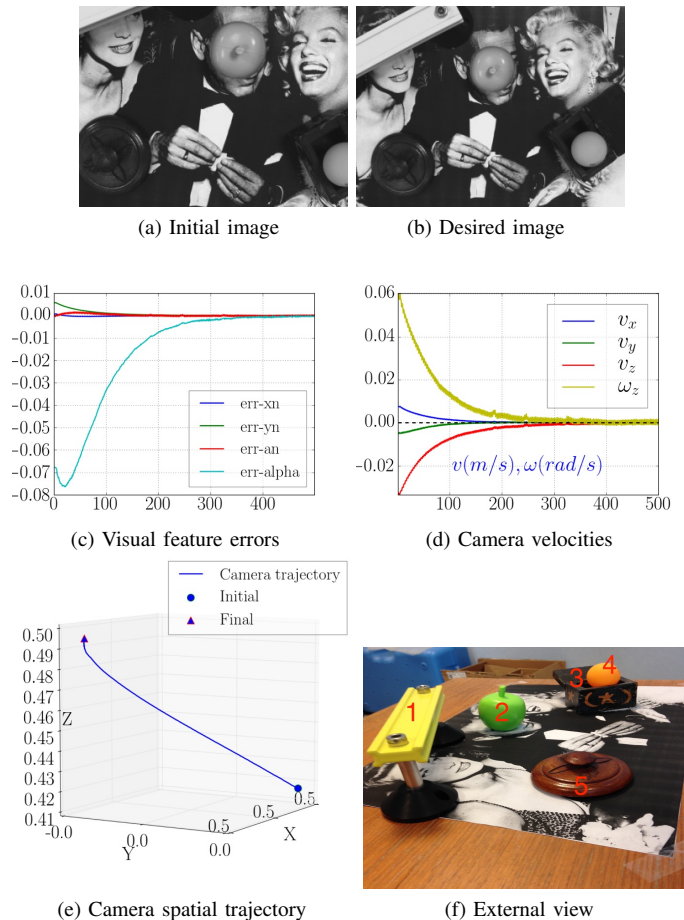


Fig. 12. 4 dof experimental results with a non planar scene (see Section V-D).

13d, it can be seen that the error in feature  $a_n$  is noisy. This feature is based on the area moment  $m_{00}$  directly related to the quantity of pixels in the image. Since the lighting conditions are not controlled, this might sometimes contribute to some noise in the features. It is also to be noted that when the interaction matrix is updated at each iteration (for the mean configuration in this case), this noise in the features sometimes make the velocities noisy as well (see Figure 13f). However, this noise does not affect the satisfactory convergence as evidenced by our results. A satisfactory Cartesian behaviour was obtained as shown in Fig 13g. The final accuracy in translations is  $[-0.05\text{mm}, 1.1\text{mm}, 0.08\text{mm}]$  and for the rotations is  $[0.18^\circ, 0.006^\circ, -0.019^\circ]$ . Let us finally note that a superior strategy would be to use the photometric moments during the beginning of the servo and to switch over to the pure luminance feature near convergence (when the error norm is below a certain lower bound). This strategy would ensure both enhanced convergence domain thanks to photometric moments and excellent accuracies at convergence thanks to luminance feature.

Let us finally note that it is possible to use a normalized intensity level in order to be robust to global lighting variations. Such a normalization can be easily obtained by computing in a first step the smallest and highest values observed in the image. This simple strategy does not modify any modelling

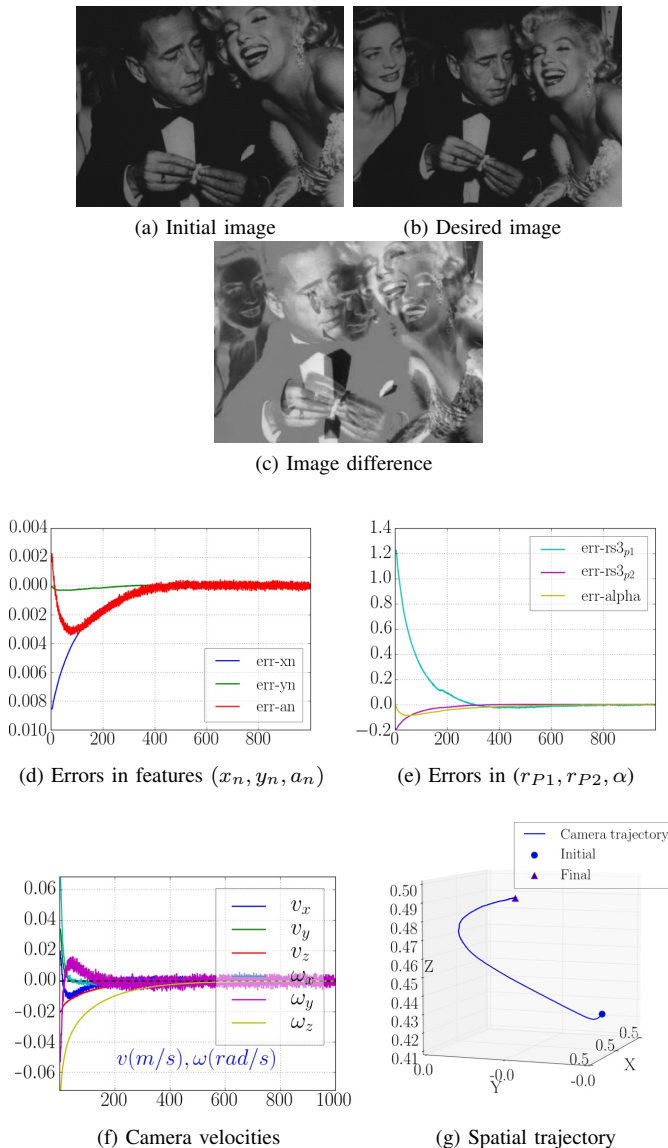


Fig. 13. WPM 6 dof experimental results (see Section V-E).

step presented in this paper as long as the parts of the scene corresponding to these extremal values do not leave the image (or new portions with higher or smaller intensities do not enter in the camera field of view), which would thus allow obtaining exactly the same results in that case. On the other hand, if the extremal values do not correspond to the same parts of the scene, the induced perturbations may cause the failure of the servoing.

## VI. CONCLUSION

This paper proposed a novel visual servoing scheme based on photometric moments, which capture the image intensities in the form of image moments. The analytical form of the interaction matrix has been derived for these new features. Visual servoing is demonstrated on scenes which do not contain a discrete set of points or monotone segmented objects. Most importantly, the proposed enhanced model takes into account the effect of the scene portions which appear and

disappear from the camera field of view during the visual servoing. Existing results based on moment invariants are then exploited to obtain visual features from the photometric moments. The control using these visual features is performant for large SCARA motions (where the images acquired during the servo have very less overlap with the desired image), with a large convergence domain in comparison to both the pure luminance feature and to features based on non-weighted moments. The proposed approach can also be used with non planar environments. This paper thus brings notable improvements over the pure luminance feature and existing moments-based VS methods.

The 6 dof control using weighted photometric moments yielded satisfactory results for small displacements to be realized. The control can be rendered suitable for large displacements if the alteration in invariance properties induced by the weighting function can be prevented. So, an important future direction of work would be about the formulation of alternate weighting strategies that preserve the invariance properties as in the non-weighted moments. This is an open and challenging problem that, once solved, would ease a complete theoretical stability and robustness analysis. Also, it is to be noted that the method will certainly fail when the shared portions between the initial and desired images are too low. Another distinction with respect to geometric approaches is that the performance depends on the image contents and hence large uniform portions with poorly texture scenes might pose issues for the servoing. Despite these obvious shortcomings, we believe that direct approaches will become more commonplace and lead to highly performant visual servoing methods.

## APPENDIX

### A. Interaction matrix of $r_{P1}$ and $r_{P2}$

In (42), on expanding the terms  $(x - x_g + x_{sh})^p$  and  $(y - y_g + y_{sh})^q$  using the binomial theorem, the shifted moments can be expressed in terms of the centred moments:

$$\tilde{\mu}_{pq} = \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} x_{sh}^k y_{sh}^l \mu_{p-k, q-l} \quad (46)$$

where

$$\mu_{pq} = \iint (x - x_g)^p (y - y_g)^q w(\mathbf{x}) I(\mathbf{x}) d\mathbf{x} dy \quad (47)$$

Differentiating (46) will yield the interaction matrix of the shifted moments.

$$\begin{aligned} \mathbf{L}_{\tilde{\mu}_{pq}} &= \mathbf{L}_{x_{sh}} \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} k x_{sh}^{k-1} y_{sh}^l \mu_{p-k, q-l} \\ &+ \mathbf{L}_{y_{sh}} \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} l x_{sh}^k y_{sh}^{l-1} \mu_{p-k, q-l} \\ &+ \mathbf{L}_{\mu_{p-k, q-l}} \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} x_{sh}^k y_{sh}^l \end{aligned} \quad (48)$$

where, from (43),

$$\begin{cases} \mathbf{L}_{x_{sh}} = \frac{1}{2} \frac{\cos \theta}{\sqrt{m_{00}}} \mathbf{L}_{m_{00}} - \sqrt{m_{00}} \sin \theta \mathbf{L}_{\theta} \\ \mathbf{L}_{y_{sh}} = \frac{1}{2} \frac{\sin \theta}{\sqrt{m_{00}}} \mathbf{L}_{m_{00}} + \sqrt{m_{00}} \cos \theta \mathbf{L}_{\theta} \end{cases} \quad (49)$$

with  $\theta = \alpha$  for shift point  $P_1$  and  $\theta = \alpha + \frac{\pi}{2}$  for shift point  $P_2$ . Further, by differentiating (47), we obtain

$$\mathbf{L}_{\mu_{pq}} = \sum_{k=0}^p \sum_{l=0}^q (-1)^r \binom{p}{k} \binom{q}{l} \left[ x_g^{p-k} y_g^{q-l} \mathbf{L}_{m_{kl}} + m_{kl} \left( (p-k)x_g^{p-k-1} y_g \mathbf{L}_{x_g} + (q-l)x_g y_g^{q-l-1} \mathbf{L}_{y_g} \right) \right] \quad (50)$$

with  $r = p + q - k - l$ . Knowing (48) and (49), the interaction matrix for any shifted moment of order  $p + q$  can be obtained. The next step is to compute  $\mathbf{L}_{\phi_1}$  and  $\mathbf{L}_{\phi_2}$  by differentiating (41). Finally, the interaction matrix  $\mathbf{L}_r$  is directly obtained by differentiating (40).

## REFERENCES

- [1] M. Bakhavatchalam, F. Chaumette, and E. Marchand, "Photometric moments: New promising candidates for visual servoing," *IEEE Int. Conf. on Robotics and Automation, ICRA'13*, Karlsruhe, Germany, May 2013, pp. 5521–5526.
- [2] M. Bakhavatchalam, F. Chaumette, and O. Tahri, "An improved modelling scheme for photometric moments with inclusion of spatial weights for visual servoing with partial appearance/disappearance," *IEEE Int. Conf. on Robotics and Automation, ICRA'15*, Seattle, WA, May 2015.
- [3] F. Chaumette and S. Hutchinson, "Visual servoing and visual tracking," *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer Berlin Heidelberg, 2008, pp. 563–583.
- [4] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. on Robotics and Automation*, 8(3):313–326, Jun 1992.
- [5] W. J. Wilson, C. Williams Hulls, and G. Bell, "Relative end-effector control using cartesian position based visual servoing," *IEEE Trans. on Robotics and Automation*, 12(5):684–696, Oct 1996.
- [6] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *IEEE Trans. on Robotics*, 20(4):713–723, Aug. 2004.
- [7] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Trans. on Robotics*, 21(6):1116–1127, Dec 2005.
- [8] O. Tahri, A. Tamtsia, Y. Mezouar, and C. Démonceaux, "Visual servoing based on shifted moments," *IEEE Trans. on Robotics*, 31(3):798–804, Apr. 2015.
- [9] S. Benhimane and E. Malis, "Homography-based 2d visual tracking and servoing," *Int. Journal of Robotics Research*, 26(7):661–676, 2007.
- [10] G. Silveira and E. Malis, "Direct visual servoing: Vision-based estimation and control using only nonmetric information," *IEEE Trans. on Robotics*, 28(4):974–980, Aug 2012.
- [11] G. Silveira, "On intensity-based nonmetric visual servoing," *IEEE Trans. on Robotics*, 30(4):1019–1026, Aug 2014.
- [12] S. Nayar, S. Nene, and H. Murase, "Subspace methods for robot vision," *IEEE Trans. on Robotics and Automation*, 12(5):750–758, Oct 1996.
- [13] K. Deguchi, "A direct interpretation of dynamic images with camera and object motions for vision guided robot control," *Int. Journal of Computer Vision*, 37(1):7–20, 2000.
- [14] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE Trans. on Robotics*, 27(4):828–834, Aug. 2011.
- [15] G. Caron, E. Marchand, and E. M. Mouaddib, "Photometric visual servoing for omnidirectional cameras," *Autonomous Robots*, 35(2):177–193, 2013.
- [16] C. Teulière and E. Marchand, "A dense and direct approach to visual servoing using depth maps," *IEEE Trans. on Robotics*, 30(5):1242–1249, Oct 2014.
- [17] V. Kallem, M. Dewan, J. Swensen, G. Hager, and N. Cowan, "Kernel-based visual servoing," *IEEE Int. Conf. on Intelligent Robots and Systems, IROS'07*, San Diego, California, Nov. 2007, pp. 1975–1980.
- [18] R. Mahony, S. Stramigioli, "A port-Hamiltonian approach to image-based visual servo control for dynamic systems," *Int. Journal of Robotics Research*, 31(11):1303–1319, 2012.
- [19] H. Wang, Y.-H. Liu, W. Chen, "Uncalibrated visual tracking control without visual velocity," *IEEE Trans. on Robotics*, 18(6):1359–1370, Nov 2010.
- [20] H. Wang, "Adaptive visual tracking for robotic systems without image-space velocity measurement", *Automatica*, 55:294–301, 2015.
- [21] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, 18(6):311–317, Jun 1975.

- [22] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, 17:185–203, 1981.
- [23] C. Collewet, E. Marchand, F. Chaumette, "Visual servoing set free from image processing," *IEEE Int. Conf. on Robotics and Automation, ICRA'08*, Pasadena, California, May 2008, pp. 81–86.
- [24] J. Swensen, V. Kallem, and N. Cowan, "Empirical characterization of convergence properties for kernel-based visual servoing," *Visual Servoing via Advanced Numerical Methods*, Lecture Notes in Control and Information Sciences, Springer London, 2010, 401:23–38.
- [25] M. Bakhavatchalam, "Utilisation of photometric moments in visual servoing," Ph.D. dissertation, University of Rennes 1, France, 2015. [Online]. Available: [https://tel.archives-ouvertes.fr/tel-01143907/file/Thesis2015\\_Manipdf](https://tel.archives-ouvertes.fr/tel-01143907/file/Thesis2015_Manipdf)
- [26] E. Malis, "Improving vision-based control using efficient second-order minimization techniques," *IEEE Int. Conf. on Robotics and Automation, ICRA'04*, 2004, pp. 1843–1848.



**Manikandan Bakhavatchalam** conducted his doctoral research at the INRIA/Irisa laboratory at Rennes and received his Ph.D degree in Computer Science from the University of Rennes, France in 2015. In 2010, he graduated from the European Masters in Advanced Robotics (EMARO) programme with double Masters degree awarded by Warsaw University of Technology, Poland and Ecole Centrale de Nantes, France. His research interests include visual servoing, deep learning methods in computer vision for biometrics and video analytics.



**Omar Tahri** was born in Fez, Morocco, in 1976. He received the Masters degree in photonics, images, and system control from the Louis Pasteur University, Strasbourg, France, in 2000 and the Ph.D. degree in computer science from the University of Rennes, Rennes, France, in March 2004. Since 2015, he has been associate professor at INSA Centre Val de Loire. His current research interests include robotics and computer vision, especially visual servoing.



**François Chaumette** (M'02, SM'09, F'13) received the Ph.D degree in computer science from the University of Rennes in 1990. Since 1990, he has been with Inria at Irisa in Rennes. His research interests include robotics and computer vision, especially visual servoing and active perception. Dr Chaumette received the 2002 King-Sun Fu Memorial Best IEEE Trans. on Robotics and Automation Paper Award. He is currently in the Editorial Board of the Int. Journal of Robotics Research, Founding Senior Editor of the IEEE Robotics and Automation Letters, and Editor of the IEEE Trans. on Robotics.