2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

A Visual Servoing Model for Generalised Cameras:
Case study of non-overlapping cameras

A.I. Comport
Laboratory I3S-CNRS
University of Nice Sophia-Antipolis
06903 Sophia Antipolis, France
Andrew.Comport@i3s.unice.fr

R. Mahony
Department of Engineering
Australian National University
ACT, O200, Australia
Robert.Mahony@anu.edu.au

F. Spindler
INRIA Rennes
Université de Rennes 1
35042 Rennes, France
Fabien.Spindler@inria.fr

***Abstract*— This paper proposes an adaptation of classical image-based visual servoing to a generalised imaging model where cameras are modelled as sets of 3D viewing rays. This new model leads to a generalised visual servoing control formalism that can be applied to any type of imaging system whether it be multi-camera, catadioptric, non-central, etc. In this paper the generalised 3D viewing cones are parameterised geometrically via Plücker line coordinates. The new visual servoing model is tested on an *a-typical* stereo-camera system with non-overlapping cameras. In this case no 3D information is available from triangulation and the system is comparable to a 2D visual servoing system with non-central ray-based control law.**

I. INTRODUCTION

The use of camera sensors to control robotic systems is an established problem that has been studied in detail for the last two decades [1], [2], [3]. In the past various cameras including perspective, stereo [4], [5], catadioptric and omni-directional cameras [6], [7] have been used to perform visual servoing [8]. The unified projection model, proposed in [9], [10], takes a first step towards a generalised camera model by classifying cameras, that view the scene through mirrors and lenses, to be grouped together with classic lens based perspective models by a two step mapping via the sphere. In this case, the perspective camera is modelled with a planar mirror. Recently advantages of performing visual servoing from spheres has also been shown in [11].

Although the unified camera model encompasses a wide class of camera models, it is limited to the class of central catadioptric sensors which requires a single effective viewpoint whereby all viewing rays pass through the same nodal point. The aim of this paper is to relax this constraint and extend the visual servoing model to encompass both central and non-central projection cameras, including multi-camera systems, in a general framework. This allows one to use a much wider class of sensors including modelling many cameras as one [12] or even conical [13] or spherical mirrors. Such sensors are particularly important in applications requiring wide angle panoramic views of the world. Furthermore, this will allow greater freedom in the design of systems because it will allow the construction of systems that don't need to be carefully designed and calibrated to ensure that the sensor(s) share a common centre of projection.

The idea of a more general camera model has been around for quite some time (note perhaps its first introduction in [14]). In the computer vision literature, the generalised camera model [15] has recently attracted much attention (see [16] and references therein). This model, which will be investigated in detail later, defines the relationship between different types of image measurements so as to unify the wide variety of camera designs. There exists a hierarchy of camera models ranging from x-slit cameras, multi-camera systems and non-central catadioptric cameras [17]. The classical perspective imaging model defines a camera as a bearing only sensor, however, when more than one camera is available or when cameras do not project centrally, different pixels sense bearings from different positions in space. In the generalised model all cameras are unified into a single sensor by modelling each pixel as sensing a cone in 3D space.

The model proposed in [15] defines the imaging cones in terms of raxels. A raxel is defined in terms of a 3D line defined by a starting point (X, Y, Z) and a direction (ϕ, θ). The imaging sensor captures light from a cone with this line as it's axis. The cone is defined by its aspect ratio and orientation (f_a, f_b, Γ). Although this model allows the radiometric quantities to vary for each pixel, this study will focus on the geometric properties and each pixel will be approximated by it's ray. An important paper in this respect is given by Pless [12] who introduced an in-depth formalisation based on defining viewing rays via Plücker lines to derive the structure from motion equations from generalised cameras. This study resulted in a comparison of the information provided by different multi-camera configurations.

A pertinent example of a generalised camera, composed of non-central projections, is a multi-camera system. Whilst the study of multi-camera systems has mainly been focused on systems with large overlap, few works have exploited the advantages of non-overlapping systems. Of course, in the former case it is possible to match features and perform 3D triangulation, however, the field of view is heavily constrained and subsequently has limited information to constrain the six degrees of freedom of the camera motion. As shown through the simulations performed in [12], the most informative configuration, in terms of the Fisher Information, is that of opposite facing cameras. One initial work on non-overlapping cameras is [18] but it was assumed that optical centers were co-located and the study thus concentrates on the case of cameras lying close together. This was extended in [19] to cameras with non-zero baseline. More recently,

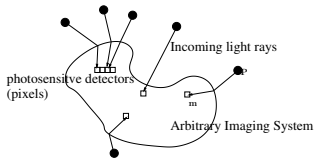

Fig. 1. (a) An arbitrary physical imaging system composed of central and non-central projection systems. The incoming light rays are projected through the imaging system onto the pixels $\mathbf{m}$. The generalised imaging system is modelled as a set of Plücker lines (or more generally raxels [15] including radiometric and optical parameters).

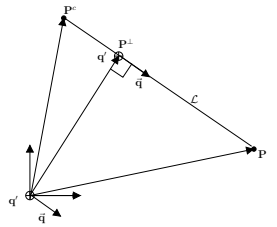

Fig. 2. A generalised camera viewing ray. The line is defined by Plücker coordinates $\mathcal{L} = (\mathbf{q}, \mathbf{q}')$ where $\mathbf{q}'$ goes into the page. $\mathbf{P}^{\perp}$ is the point on the line closest the origin. $\mathbf{P}^{c}$ is a point on the line that is rigidly linked to the generalised camera. $\mathbf{P}$ is a point in the 3D scene.

the case of non-overlapping cameras has also been studied in [20], [21] to perform structure and motion estimation. There is also a conceptual relation to earlier studies on hand-eye calibration such as [22] (as will be used in the experiments) since they allow to determine the relationship between non-overlapping cameras via known movement.

The remainder of the paper is presented as follows. Section II-A defines the Plücker coordinates of viewing rays necessary for the remainder of the paper. Section III outlines the visual servoing approach and integrates the use of generalised viewing rays into a control scheme. Section IV looks at the case study of the generalised camera model applied to two non-overlapping cameras.

II. Generalised camera model

The generalised camera model is depicted in Figure 1. A point $\mathbf{P} \in \mathbb{R}^3$ is imaged at a pixel $\mathbf{m} = (u, v)$ that is found by following the ray through the physical system. In the generalised model, each ray leaving the viewing surface can be modeled by a 3D line. Several groups have calibrated these cameras through non-conventional models such that each pixel is individually mapped to a particular viewing ray through lookup tables. The interested reader may refer to [15], [23].

A. 3D Viewing Rays and Plücker lines

In order to formalise a generalised camera model, a parameterisation for the viewing ray of each pixel will be given. A line can be defined by the *join* of two points or the intersection of two planes. Many parameterisations exist ranging from the null space and span representation, to Plücker matrices and Plücker vectors (see [24] for a brief survey). In this work Plücker vectors will be employed as in [12] and a short summary is given here.

Let $\mathbf{P}_1$ and $\mathbf{P}_2 \in \mathbb{R}^3$ be two 3D points defining a line and $\overline{\mathbf{P}}_1 = (X_1, Y_1, Z_1, W_1)^\top$ and $\overline{\mathbf{P}}_2 = (X_2, Y_2, Z_2, W_2)^\top \in \mathbb{RP}^3$ be their homogeneous coordinates. The line joining the two points is given as the Plücker matrix [24]:

$$\mathbf{L} = \overline{\mathbf{P}}_1 \overline{\mathbf{P}}_2^\top - \overline{\mathbf{P}}_2 \overline{\mathbf{P}}_1^\top. \tag{1}$$

The Plücker coordinates are subsequently the 6 non-zero independent elements of this 4×4 skew-symmetric matrix given as:

$$\mathcal{L} = \begin{pmatrix} \mathbf{L}_{41} \\ \mathbf{L}_{42} \\ \mathbf{L}_{43} \\ \mathbf{L}_{32} \\ -\mathbf{L}_{31} \\ \mathbf{L}_{21} \end{pmatrix} = \begin{pmatrix} W_1 X_2 - X_1 W_2 \\ W_1 Y_2 - Y_1 W_2 \\ W_1 Z_2 - Z_1 W_2 \\ Z_1 Y_2 - Y_1 Z_2 \\ X_1 Z_2 - Z_1 X_2 \\ Y_1 X_2 - X_1 Y_2 \end{pmatrix} \in \mathbb{RP}^5, \tag{2}$$

giving the Plücker coordinates as $\mathcal{L} = (\mathbf{q}, \mathbf{q}')^\top$.

With the normalised homogeneous coordinate $W_1 = W_2 = 1$, (2) simplifies to give $\mathbf{q} = \mathbf{P}_2 - \mathbf{P}_1$ as the direction vector (of any length) and $\mathbf{q}' = \mathbf{P}_1 \times \mathbf{P}_2$ is the moment vector of the line. Geometrically, the moment vector is perpendicular to the plane containing $\mathcal{L}$ and the origin, and its magnitude is twice the area of the triangle formed by the two points and the origin. Substituting for $\mathbf{P}_1$ in $\mathbf{q}'$, it can be easily seen that the moment vector is orthogonal to any point on the line $\mathbf{P}$ and the vector $\mathbf{q}$ so that:

$$\mathbf{q}' = \mathbf{q} \times \mathbf{P}. \tag{3}$$

See Figure 2 for a visualisation.

The line $\mathcal{L}$ has 4 degrees of freedom which can be determined as follows. The Plücker line defined in (2) has 6 parameters, however, they are only unique up to scale (only the 5 ratios are significant). Secondly, by definition the moment vector is perpendicular to every displacement along the line, so:

$$\mathbf{q}^T \mathbf{q}' = 0, \tag{4}$$

which is quadratic in the Plücker line coordinates and can be obtained by evaluating $\det(\mathbf{L}) = 0$. This leaves a minimal representation of 4 parameters.

In this paper the direction vector, $\mathbf{q}$, is normalised to be a unit vector so as to define the scale of the remaining homogeneous parameters:

$$\vec{\mathbf{q}} = \tfrac{\mathbf{q}}{|\mathbf{q}|}, \quad \vec{\mathbf{q}} \in S^2. \tag{5}$$

In this case, the point $\mathbf{P} = \vec{\mathbf{q}} \times \mathbf{q}'$ is the point on the line closest to the origin which is denoted here $\mathbf{P}^{\perp}$.

III. Visual servoing

A. Visual Servoing Model

In classic image-based control [1], [2], many different types of features $\mathbf{s}$ have been used to perform visual servoing. Interesting features range from sets of points to lines,

ellipses, cylinders, distances to 3D models and moments using various types of optical systems. In all these cases the 2D image-sensor measurements $\mathbf{m}$ belong to rays passing through a common nodal point.

In general, a vision based control scheme aims to define a task function for the robot as [8]:

$$\mathbf{e}(t) = \mathbf{s}(\mathbf{m}(t), \mathbf{a}) - \mathbf{s}^*, \tag{6}$$

where the vector $\mathbf{m}(t)$ is usually a set of image measurements in pixel coordinates, $\mathbf{a}$ is a vector containing any *a priori* knowledge about the system such as intrinsic or extrinsic camera calibration parameters or a 3D object model and where $\mathbf{s}(\mathbf{m}(t), \mathbf{a})$ is a vector of n visual features that has been constructed from the image measurements and may take into account the prior knowledge such as pixel to metric conversion. $\mathbf{s}^*$ is a corresponding vector of features containing their positions in the desired image.

The design of the control scheme then requires relating the movement of the visual measurements in the image to the movement of the robot. A first step is to determine the relationship between the time variation of $\mathbf{s}$ and the camera. The instantaneous spatial velocity of the camera $\mathbf{x} \in \mathbb{R}^6$ is parameterised by the Lie algebra $\mathfrak{se}(3)$ as:

$$\mathbf{x} = (\boldsymbol{\upsilon}, \boldsymbol{\omega})^\top \in \mathfrak{se}(3), \tag{7}$$

where $\boldsymbol{\upsilon}$ and $\boldsymbol{\omega}$ are the instantaneous linear and angular velocities respectively. This velocity twist is related to a displacement via integration as:

$$\mathbf{T}(\mathbf{x}) = \exp\left(\begin{bmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{\upsilon} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right), \tag{8}$$

where $\exp$ is the exponential map, $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in \mathbb{S}E(3)$ is the homogeneous matrix of describing the camera displacement, $\mathbf{R} \in \mathbb{S}O(3)$ is a rotation matrix such that $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$ and $\det(\mathbf{R}) = 1$, $\mathbf{t} \in \mathbb{R}^3$ is the translation vector, $[.]_\times$ represents the skew symmetric matrix operator and where the integration period is taken as $\Delta t = 1$.

The kinematics of the task function (6) for a eye-in-hand system are then given as:

$$\dot{\mathbf{e}} = \mathbf{L}\mathbf{x}, \tag{9}$$

where $\mathbf{L}$ is the *interaction matrix* related to $\mathbf{s}$.

If the robot is controlled by velocity then $\mathbf{x}$ can be taken as the input to the robot controller. If an exponential decrease of the error is imposed then $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ and the control law is given by:

$$\mathbf{x} = -\lambda\widehat{\mathbf{L}}^+\mathbf{e}, \tag{10}$$

where the hat operator signifies that the interaction matrix is computed from uncertain measurements and therefore it is only an approximation. $\mathbf{L}^+ = (\mathbf{L}^\top\mathbf{L})\mathbf{L}^\top \in \mathbb{R}^{6\times n}$ is the pseudo inverse of $\mathbf{L}$ which must be of full rank 6.

In summary, the classic model is quite general in that any type of feature $\mathbf{s}$ may be constructed from both $\mathbf{m}$ and the *a-priori* parameters $\mathbf{a}$. It is also worth noting that multi-camera systems have been classically modelled by forming the task error directly in each camera's reference frame and not via generalised coordinates as will be shown in the next section.

B. Generalised Visual Servoing

In this paper, the generalised camera model is used to construct a visual servoing control law in which the sensor measures a set of 3D viewing rays $\boldsymbol{\mathcal{L}} = (\mathcal{L}_1^\top, \ldots, \mathcal{L}_n^\top)^\top$ with n the number of rays (and not just 2D pixel coordinates $\mathbf{m}$ as does a classic camera sensor). In this way (2) defines a broad group of possible physical measurements and represents a general class of cameras including non-central projection cameras. This encompasses non-central projection cameras including multi-camera systems. An important result of this is that it amounts to considering that the sensor measure's noisy lines in space whereby the line coordinates encompass all measurement errors including both classical 2D measurement and extrinsic calibration errors.

Given this general model, the rays can be constructed from any type of sensor and the visual servoing task (6) becomes:

$$\mathbf{e}(t) = \mathbf{s}(\boldsymbol{\mathcal{L}}(\mathbf{a}, t), \mathbf{b}) - \mathbf{s}^*, \tag{11}$$

and where $\mathbf{a}$ is now redefined as a vector of general camera parameters and $\mathbf{b}$ is a vector containing any *a priori* knowledge about the system such as a 3D object model, etc.

1) **Perspective camera:** In this case the viewing rays of the general model can be shown to depend only on a classic pinhole model as $\mathcal{L}(\mathbf{m}(t), \boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is the vector of the intrinsic camera parameters required to determine the normalised rays. Under the generalised model a pixel coordinate forms a 3D ray $\mathcal{L}$ that passes through the image plane at the point $\mathbf{m}$ and intersects with the optical centre of the camera. If the pinhole camera is calibrated with the calibration matrix $\mathbf{K}(\boldsymbol{\xi}) \in \mathbb{R}^{3\times3}$, then a normalised point $\overline{\mathbf{p}} = (x, y, 1)$ may be determined from the image coordinates. In this case the moment vector is $\mathbf{q}' = \mathbf{0}$ since the origin, $\mathbf{P} = (0, 0, 0)$, belongs to the ray. The 3D ray is subsequently defined by the Plücker vector $\mathcal{L} = (\frac{\mathbf{K}\overline{\mathbf{m}}}{|\mathbf{K}\overline{\mathbf{m}}|}, \mathbf{0})$.

2) **Multi-camera system:** On the other hand, if the camera has multiple centres of projection, the moment vector is non-null. For example, take a multi-camera system as a generalised non-central projection camera. In this case the rays depend on each camera's pixels $\mathcal{M} = (\mathbf{m}_1, \ldots, \mathbf{m}_n)$ and the cameras' intrinsic and extrinsic parameters $\mathbf{a} = (\boldsymbol{\xi}^c, \mathbf{x}^c)$. In this case each camera c has a moment vector obtained from (15) as $\mathbf{q}' = \vec{\mathbf{q}} \times \mathbf{P}^\mathbf{c}$, where $\mathbf{P}^c = -\mathbf{R}^{e\top}\mathbf{t}^e$ is the origin of the camera seen in the generalised coordinate system with $(\mathbf{R}^e, \mathbf{t}^e)$ being the extrinsic camera rotation and translation parameters. The direction vector is $\vec{\mathbf{q}} = \mathbf{R}^{e\top}\vec{\mathbf{q}} = \mathbf{R}^{e\top}\frac{\mathbf{K}\overline{\mathbf{m}}}{|\mathbf{K}\overline{\mathbf{m}}|}$.

3) **Generalised-camera system:** Consider now, a generalised-camera system that *directly* measures a set of viewing rays $\boldsymbol{\mathcal{L}}(\mathbf{q}(t), \mathbf{q}'(t))$ defined in a generalised reference frame. It is attached to an eye-in-hand visual servoing system that moves wrt. an inertial coordinate system. Viewing rays are directly used to define the task function (higher order features $\mathbf{s}(\boldsymbol{\mathcal{L}})$ are not considered here):

$$\mathbf{e}(t) = \boldsymbol{\mathcal{L}}(t) - \boldsymbol{\mathcal{L}}^*, \tag{12}$$

The set of all viewing rays are related to movement of the camera via the analytical relationship given in section II-A. Each direction vector is related to a 3D world point as:

$$\mathbf{q}_i = \mathbf{P}_i^w(t) - \mathbf{P}_i^c, \tag{13}$$

where $\mathbf{P}_i^w$ is the coordinate of a point in the inertial world coordinate system and $\mathbf{P}_i^c$ is the point which is rigidly fixed to the generalised camera (for a perspective camera it is the optical center). To simplify the discussion, a-priori knowledge of $\mathbf{P}^c$ is considered available, however, it will be shown in Subsection III-B.4 that is not required in a generalised camera system.

From (13), the normalised unit direction vector is:

$$\vec{\mathbf{q}}(t) = \frac{\mathbf{q}(t)}{|\mathbf{q}(t)|} = \frac{\mathbf{P}^w(t) - \mathbf{P}^c}{|\mathbf{q}(t)|}, \tag{14}$$

where the ray index is assumed implicit.

The moment vector defining each view ray is subsequently:

$$\mathbf{q}'(t) = \vec{\mathbf{q}}(t) \times \mathbf{P}^c. \tag{15}$$

Taking first the time derivative of the direction vector equation (14) gives:

$$\dot{\vec{\mathbf{q}}} = |\mathbf{q}|^{-1}\dot{\mathbf{q}} + \frac{d}{dt}(|\mathbf{q}|^{-1})\mathbf{q}, \tag{16}$$

where $\dot{\mathbf{q}} = \dot{\mathbf{P}}^w$ giving well known motion for a point:

$$\dot{\mathbf{q}} = -[\boldsymbol{\omega}]_\times \mathbf{q} - (\boldsymbol{\upsilon} + [\boldsymbol{\omega}]_\times \mathbf{P}^c) \tag{17}$$

where both $\boldsymbol{\upsilon}$ and $\boldsymbol{\omega}$ form the velocity twist of the generalised reference frame wrt. the world frame.

The second term of (16) is developed and simplified as:

$$\begin{aligned}
\tfrac{d}{dt}\left(|\mathbf{q}|^{-1}\right) &= \tfrac{d}{dt}\left(\left(\mathbf{q}^\top\mathbf{q}\right)^{-\frac{1}{2}}\right), \\
&= -\tfrac{1}{2}\tfrac{d}{dt}\left(\mathbf{q}^\top\mathbf{q}\right)\left(\mathbf{q}^\top\mathbf{q}\right)^{-\frac{3}{2}}, \\
&= -\tfrac{1}{2}\left(\dot{\mathbf{q}}^\top\mathbf{q}\right)\left(\mathbf{q}^\top\mathbf{q}\right)^{-\frac{3}{2}} - \tfrac{1}{2}\left(\mathbf{q}^\top\dot{\mathbf{q}}\right)\left(\mathbf{q}^\top\mathbf{q}\right)^{-\frac{3}{2}}, \\
&= -\left(|\mathbf{q}|^3\right)^{-1}\left(\mathbf{q}^\top\dot{\mathbf{q}}\right).
\end{aligned} \tag{18}$$

Substituting (17) and (18) into (16) gives the dynamic equation for the unit direction vector as:

$$\begin{aligned}
\dot{\vec{\mathbf{q}}} &= \tfrac{1}{|\mathbf{q}|}\dot{\mathbf{q}} - \tfrac{1}{|\mathbf{q}|^3}\mathbf{q}^\top\dot{\mathbf{q}}\mathbf{q}, \\
&= -\tfrac{1}{|\mathbf{q}|}[\boldsymbol{\omega}]_\times\mathbf{q} - \tfrac{1}{|\mathbf{q}|}\pi_{\vec{\mathbf{q}}}\left(\boldsymbol{\upsilon} + [\boldsymbol{\omega}]_\times\mathbf{P}^c\right),
\end{aligned} \tag{19}$$

where $\pi_{\vec{\mathbf{q}}} = (\mathbf{I} - \vec{\mathbf{q}}\vec{\mathbf{q}}^\top)$ is an orthogonal projection $\pi_{\vec{\mathbf{q}}} : \mathbb{R}^3 \rightarrow T_{\vec{\mathbf{q}}}\mathbb{S}^2$ onto the tangent space of the sphere $\mathbb{S}^2$ at the point $\vec{\mathbf{q}} \in \mathbb{S}^2$.

Equation (19) is equivalent to the adjoint transformation of the spherical projection of a point in a classic perspective camera in which the unknown $\mathbf{P}^w$ can be eliminated from (19) as:

$$\dot{\vec{\mathbf{q}}} = [\vec{\mathbf{q}}]_\times\boldsymbol{\omega} + \frac{1}{|\mathbf{q}|}\pi_{\vec{\mathbf{q}}}\left(\boldsymbol{\upsilon} + [\boldsymbol{\omega}]_\times\mathbf{P}^c\right), \tag{20}$$

where the skew symmetric property $[\mathbf{a}]_\times\mathbf{b} = -[\mathbf{b}]_\times\mathbf{a}$ has been used.

Taking then, the time derivative of equation (15) gives simply:

$$\dot{\mathbf{q}}' = \dot{\vec{\mathbf{q}}} \times \mathbf{P}^c \tag{21}$$

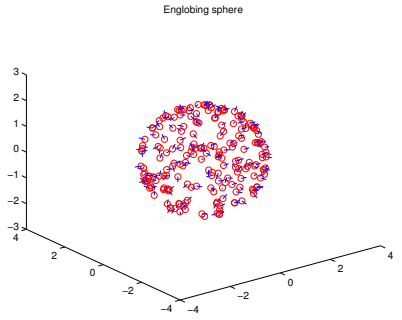

Fig. 3. (a) A simulated generalised camera (used in Section IV-A) which is 'calibrated' such that a sphere encloses all viewing rays. This is simpler since there is only one "base-line" parameter (the sphere radius).

4) **Generalised parametrisation:** Here it is assumed that a mapping from sensor space to the generalised coordinate space has been achieved by a *general* calibration model [15], [23] whereby the viewing rays contain all necessary information. In that case no other information is required such as calibration parameters. This means that the center of projection $\mathbf{P}^c$ used previously is unknown. This system can be defined simply by a sphere that englobes all the viewing rays (see Figure 3).

From each $\mathbf{q}$ and $\mathbf{q}'$ it is possible to determine the point of the line that is closest to the origin of the generalised camera system as:

$$\mathbf{P}^\perp(t) = \vec{\mathbf{q}}(t) \times \mathbf{q}'(t), \tag{22}$$

The variation of $\dot{\mathbf{P}}^\perp$ is then approximated as a function of its gradient:

$$\dot{\mathbf{P}}^\perp(t) \approx \mathbf{P}(t)^\perp - \mathbf{P}^{\perp *} \tag{23}$$

which is determined directly from the current $\boldsymbol{\mathcal{L}}(t)$ and desired $\boldsymbol{\mathcal{L}}^*$ viewing rays. It's maximum variation is bounded by $\mathbf{P}^c$ and it is assumed that the control law will be chosen to maintain this assumption.

In that case (13) is:

$$\mathbf{q}^\perp = \mathbf{P}^w(t) - \mathbf{P}^\perp(t). \tag{24}$$

Following through the same derivation in the previous section leads to the dynamic equation for the unit direction vector as:

$$\dot{\vec{\mathbf{q}}} = \frac{1}{|\mathbf{q}^\perp|}\pi_{\vec{\mathbf{q}}}(\dot{\mathbf{P}}^w + \dot{\mathbf{P}}^\perp). \tag{25}$$

Since equation (17) requires $\mathbf{P}^w$, it can be substituted using (13) and (14) by:

$$\mathbf{P}^w = |\mathbf{q}^\perp|\vec{\mathbf{q}} + \mathbf{P}^\perp, \tag{26}$$

into (25) to give:

$$\dot{\vec{\mathbf{q}}} = [\vec{\mathbf{q}}]_\times\boldsymbol{\omega} - \frac{1}{|\mathbf{q}^\perp|}\pi_{\vec{\mathbf{q}}}\left(\boldsymbol{\upsilon} + [[\vec{\mathbf{q}}]_\times\mathbf{q}']_\times\boldsymbol{\omega} + \dot{\mathbf{P}}^\perp\right). \tag{27}$$

If the moment vector is evaluated with the current definition using (22) to give:

$$\dot{\mathbf{q}}' = \vec{\mathbf{q}} \times \dot{\mathbf{P}}^\perp - [[\mathbf{q}]_\times\mathbf{q}']_\times\dot{\vec{\mathbf{q}}}. \tag{28}$$

$$\mathbf{L}_{\mathcal{L}} = \begin{bmatrix} -\frac{1}{|\mathbf{q}^{\perp}|}\pi_{\vec{\mathbf{q}}} & [\vec{\mathbf{q}}]_{\times} + \frac{1}{|\mathbf{q}^{\perp}|}\pi_{\vec{\mathbf{q}}}[[\vec{\mathbf{q}}]_{\times}\mathbf{q}']_{\times} \\ \frac{1}{|\mathbf{q}^{\perp}|}[[\vec{\mathbf{q}}]_{\times}\mathbf{q}']_{\times}\pi_{\vec{\mathbf{q}}} & -\left([[\vec{\mathbf{q}}]_{\times}\mathbf{q}']_{\times}\right)\left([\vec{\mathbf{q}}]_{\times} + \frac{1}{|\mathbf{q}^{\perp}|}\pi_{\vec{\mathbf{q}}}[[\vec{\mathbf{q}}]_{\times}\mathbf{q}']_{\times}\right) \end{bmatrix} \tag{28}$$

The resulting equations (27) and (28) relate the velocity of the 3D point to the spatial velocity of a viewing ray as:

$$\dot{\mathcal{L}} = \mathbf{L}_{\mathcal{L}}\mathbf{x} + \mathbf{B}, \tag{29}$$

where $\mathbf{x}$ is the kinematic screw between the generalised camera frame and the world and $\mathbf{L}_{\mathcal{L}}$ is given in (28). $\mathbf{B} = \left(-\frac{1}{|\mathbf{q}^{\perp}|}\pi_{\vec{\mathbf{q}}}\dot{\mathbf{P}}^{\perp}, \vec{\mathbf{q}} \times \dot{\mathbf{P}}^{\perp}\right)^{\top}$ is the bias term related due to $\mathbf{P}^{\perp}$.

Even if the interaction matrix is rank 2 (since $\dot{\mathbf{q}}'$ is a linear combination of $\mathbf{q}$), using both measurement vectors provides increased robustness with respect to measurement noise. Notably, if there is more noise in $\vec{\mathbf{q}}$ than $\mathbf{q}'$ then the moment vector helps minimise this (see Figure 4(b)). Similarly to classical visual servoing, a rough estimate of the depth $|\mathbf{q}^{\perp}|$ of the 3D point to $\mathbf{P}^{\perp}$ and is required in practice.

In order to have full rank 6 to control 6dof it is necessary to have at least three non-degenerate and non-singular points in a single view configuration [25]. In a stereo-perspective case it is necessary to avoid that the 3D point lies on the baseline [24]. Here the control law is built by stacking multiple viewing ray interaction matrices with more than three non-coplanar points in an appropriate configuration to avoid singularities or ambiguous global minima.

The stability of the system can be analysed by considering the $l^2 - norm$ candidate Lyapunov function $\mathcal{V} = |\sum_{i=1}^{n} e_i|$ which is derived as:

$$\begin{aligned} \dot{\mathcal{V}} &= {\mathbf{e}_{\mathbf{B}}}^{\top}\dot{\mathbf{e}}_{\mathbf{B}} \\ &= -\lambda {\mathbf{e}_{\mathbf{B}}}^{\top}\mathbf{L}_{\mathbf{e}}\widehat{\mathbf{L}}_{\mathbf{e}}^{+}\mathbf{e}_{\mathbf{B}}, \end{aligned} \tag{31}$$

where $\mathbf{e}_{\mathbf{B}} = \mathbf{e} - \mathbf{B}$ and from which the global asymptotic stability is ensured if the following condition is respected:

$$\mathbf{L}_{\mathbf{e}}\widehat{\mathbf{L}}_{\mathbf{e}}^{+} > 0. \tag{32}$$

IV. Results

In order to test this model a series of simulations were performed followed by validation on a real robotic platform. An *eye-in-hand* visual servoing experiment is considered.

A. Simulation

Several simulations were perfomed to test the generalised camera visual servoing configuration (see Figure 3). In the simulation an arbitrary number of centers of projection and an arbitrary number of points were generated. In the results shown in Figure 4, 200 projection centers were considered with 10 lines projecting onto each center. The initial general camera pose was set to $(0.5, 0.1, 2.5, 280^o, 52^o, 0^o)$ and the final to $(0, 0, 1, 0^o, 0^o, 0^o)$. Noise was added onto both $\mathbf{q}$ and $\mathbf{q}'$ and the effect is compared and as expected the added task vector for $\mathbf{q}'$ improves when there is a large amount of noise on $\mathbf{q}$. Even with a large rotation of 180^o around the x-axis, the control performs quite nicely. Of course this assumes that the points are visibile over a large domain which is not easily the case in practice as will be seen in the next experiment.

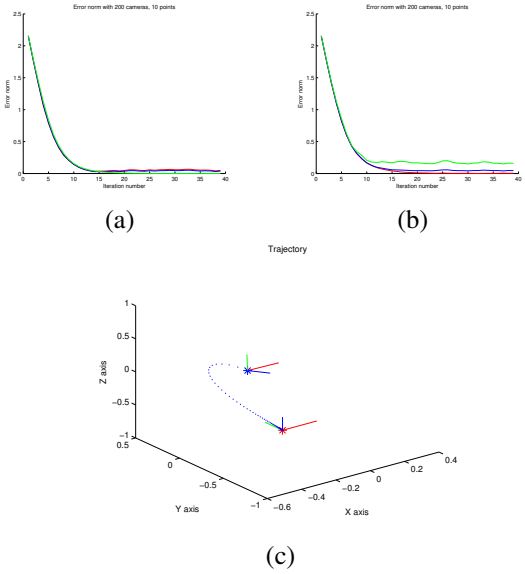

Fig. 4. A visual servoing simulation for a generalised camera configuration using 200 randomly generated centers of projection on a sphere with radius $1m$ with 10 random points in each. A large rotation of 180^o around the x axis was made. Red is $\mathbf{q}$, blue is $\boldsymbol{\mathcal{L}}$ and yellow $\mathbf{q}'$. (a) The error norm with $\sigma = 0.05$ Gaussian noise on $\mathbf{q}$,(b) The error norm with the same noise on $\mathbf{q}'$. (c) The trajectory in 3D with the red the desired position, blue the trajectory, and the blue star the initial camera position.

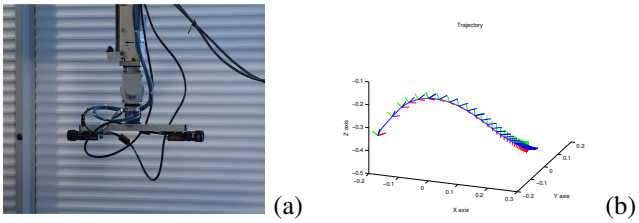

Fig. 5. (a) The non-overlapping stereo eye-in-hand configuration attached to INRIA's 6dof. Afma 6 robot. (b) A 3D trajectory of the robot performing visual servoing from (12). A video of the experiment is available accompanying this paper.

B. Non-overlapping visual servoing

Since both cameras are not viewing the same 3D points, conventional extrinsic camera calibration was not possible. It was necessary to calibrate each camera individually with respect to a common coordinate system. In this case the cameras were mounted on the robot (see Figure 5) and the robot was used as the common coordinate frame. The robot makes a series of automatically planned movements with a camera rigidly mounted at the gripper. At the end of each move, feature coordinates are extracted and camera extrinsic calibration is performed using the technique given in [22].

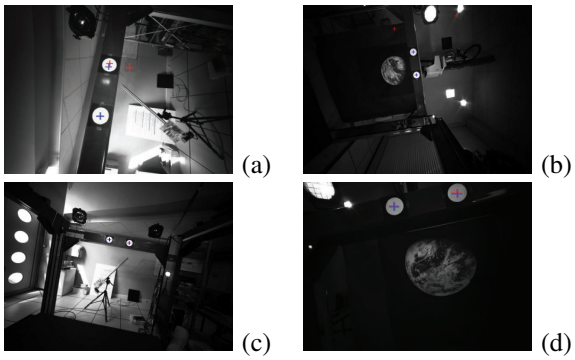

Fig. 6. A visual servoing experiment for a non-overlapping stereo eye-in-hand configuration using 2 viewing rays in each camera. The red crosses display the desired image positions and the blue crosses the current point match. (a) Camera 1's initial image. (b) Camera 2's initial image. (c) Camera 1's 149th image. (d) Camera 2's 149th image.

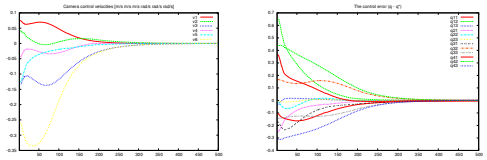

Fig. 7. The non-overlapping stereo visual servoing experiment (a) The camera control velocities. (b) The error of the control vector.

To test the limits of the system a task involving a near 90deg rotation around the z axis, along with a considerable translation was considered. The initial and desired images are shown in Figure 6. It can be seen that two points have been used in each image marked by blue crosses. The desired positions, marked in red, have been determined from a desired robot pose and model-based pose estimation has been used to determine the initial position. As can be seen in Figure 7 (a) and (b), an exponential decrease in the overall error is obtained and smooth velocities are used to control the robot. Furthermore a smooth trajectory in both rotation and translation is obtained in 3D shown in Figure 5(b).

V. CONCLUSIONS AND FUTURE WORKS

In conclusion, this paper presented an alternative visual servoing model based on the concept of a generalised camera. This was achieved through the use of Plücker line coordinates to geometrically model general viewing rays in 3D. This has allowed to develop a visual servoing control law that has been tested on a non-overlapping stereo configuration and shown to give very satisfactory results. Future work will be dedicated to the investigation of the use of this general formalism to develop interesting higher level control features.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the INRIA Lagadic team for use of their Afma 6 robot and software [26]. This research was supported by the CNRS and by the Australian Research Council through discovery grant DP0880509, "Image-based teleoperation of semi-autonomous robotic vehicles".

REFERENCES

[1] L. Weiss, A. Sanderson, and C. Neuman, "Dynamic sensor-based control of robots with visual feedback," *IEEE Journal of Robotics and Automation*, vol. 3, no. 5, pp. 404–417, October 1987.
[2] J. Feddema and O. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 5, pp. 691–700, October 1989.
[3] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, June 1992.
[4] G. Hager, W. Chang, and A. Morse, "Robot feedback control based on stereo vision: Towards calibration-free hand-eye coordination," *IEEE Control Systems Magazine*, vol. 15, no. 1, pp. 30–39, 1995.
[5] E. Malis, G. Morel, and F. Chaumette, "Robot control from disparate multiple sensors," *International Journal of Robotics Research*, vol. 20, no. 5, pp. 364–378, May 2001.
[6] Y. Mezouar, H. Hadj Abdelkader, P. Martinet, and F. Chaumette, "Visual servoing from 3d straight lines with central catadioptric cameras," in *OmniVis Workshop*, Prague, Czech Republic, May 2004.
[7] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke, "Generic decoupled image-based visual servoing for cameras obeying the unified projection model," in *IEEE International Conference on Robotics and Automation*, Kobe, Japan, May 12-17 2009, pp. 1116–1121.
[8] F. Chaumette and S. Hutchinson, "Visual servo control, part ii: Advanced approaches," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, pp. 109–118, March 2007.
[9] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems," in *Proceedings 6th European Conference on Computer Vision*, Dublin, Ireland, 2000, pp. 445–462.
[10] ——, "Catadioptric projective geometry," *International Journal of Computer Vision*, vol. 45, pp. 223–243, 2002.
[11] P. I. Corke, "Spherical image-based visual servo and structure estimation," in *ICRA*, 2010, pp. 5550–5555.
[12] R. Pless, "Using many cameras as one," in *International Conference on Computer Vision and Pattern Recognition*, vol. 2, Madison, Wisconsin, USA, 16-22 June 2003, pp. 587–593.
[13] C. Cauchois, E. Brassart, L. Delahoche, and A. Clerentin, "3d localization with conical vision," in *OmniVis Workshop*, 2003.
[14] K. Gremban, C. Thorpe, and T. Kanade, "Geometric camera calibration using systems of linear equations," in *IEEE Int. Conference on Robotics and Automation*, vol. 1, April 1988, pp. 562–567.
[15] M. Grossberg and S. Nayar, "A general imaging model and a method for finding its parameters," in *IEEE Int. Conference on Computer Vision*, vol. 2, Vancouver, Canada, 7-14, July 2001, pp. 108–115.
[16] H. Li, R. I. Hartley, and J.-H. Kim, "A linear approach to motion estimation using generalized camera models," in *CVPR*, 2008.
[17] P. Sturm, "Multi-view geometry for general camera models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California*, vol. 1, jun 2005, pp. 206–212.
[18] Y. Caspi and M. Irani, "Aligning non-overlapping sequences," *International Journal of Computer Vision*, vol. 48, no. 1, pp. 39–51, 2002.
[19] F. Dornaika and C. Chung, "Stereo geometry from 3d ego-motion streams," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 33, no. 2, pp. 308–323, April 2003.
[20] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Generic and real time structure from motion," in *18th British Machine Vision Conference*, Warwick, United Kingdom, September 2007.
[21] B. Clipp, J. Kim, J. Frahm, and R. Pollefeys, M. Hartley, "Robust 6dof motion estimation for non-overlapping, multi-camera systems," in *IEEE Workshop on Applications of Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 1–8.
[22] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, June 1989.
[23] P. Sturm and S. Ramalingam, "A generic concept for camera calibration," in *Proceedings of the European Conference on Computer Vision, Prague, Czech Republic*, vol. 2. Springer, May 2004, pp. 1–13.
[24] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
[25] H. Michel and P. Rives, "Singularities in the determination of the situation of a robot effector from the perspective view of three points," in *INRIA Research Report, No. 1850*, Feburary 1993.
[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 & Automation Magazine*, vol. 12, no. 4, Dec. 2005.