Weakly Calibrated Stereoscopic Visual Servoing for Laser Steering: Application to Phonomicrosurgery

Brahim Tamadazte and Nicolas Andreff

Abstract—This paper deals with the study of a weakly calibrated multiview visual servoing control law for micro-robotic laser phonomicrosurgery of the vocal folds. It consists of the development of an endoluminal surgery system for laser ablation and resection of cancerous tissues. More specifically, this paper focuses on the part concerning the control of the laser spot displacement during surgical interventions. To perform this, a visual control law based on trifocal geometry is designed using two cameras and a laser source (virtual camera). The method is validated on a realistic testbench and the straight point-to-point trajectories are demonstrated.

I. INTRODUCTION

The direct diagnosis of the larynx and the trachea are essential in the Otolaryngology field. This is possible using multiple existing endoscopic vision systems. The first endoscopic system for larynx inspection is probably the glottoscope developed by [1], in 1828 (Fig. 1(a)). Since then, these systems have largely evolved towards more efficient inspection and surgery intervention devices. The most successful protocol for surgery of vocal folds, widely used in hospitals is certainly the suspension laryngoscopic microrobotic laser phonomicrosurgery of the vocal folds. This consists of an endoscope which is adjustable laryngoscope for microrobotic laser phonomicrosurgery of the vocal folds (Fig. 1(b)). This disposition involves the projection, in straight line, of the laser beam on the soft tissue through the patient’s mouth and the laryngoscope tunnel. This causes a very uncomfortable position for the patient with an extreme extension of the neck, that makes it painful several days after the operation [3]. Apart from this, the requirement of dexterity and considerable expertise of the surgeon to perform an intervention of the vocal cords are highly required. This is true, among other, because the laser displacements are performed in open-loop control.

In our work, we propose a new concept of a more intuitive adjustable laryngoscope for microrobotic laser phonomicrosurgery of the vocal cords. This consists of an endoscope which is equipped with a two degrees-of-freedom (dof) micromirror, a fixed mirror, two image bundles, a lighting source, and a laser source (Fig. 1(c)). This paper focuses specifically on the design of a closed-loop control strategy to achieve automatic laser partial ablation or resection of local cancerous tissues. Using a closed-loop visual feedback control law, commonly called visual servoing, is justified by its robustness to calibration errors (camera and robot) and to changes in the robot’s environment [4]. The proposed visual servoing control law is based on the use of the multiview geometrical constraints. Indeed, with two real cameras and one virtual camera (i.e., if we consider the laser beam as a virtual camera), it is possible to build a trifocal geometry system. Thereby, using the trifocal constraint, it is possible to design an original visual servoing control law.

Visual servoing for visible laser control has been widely studied in the literature for many years. This covers several applications in industrial [5], [6], mobile [7], [8] or surgical robotics [9]. The laser scanner or pointer is generally used in two configurations: eye-to-hand [10] or eye-in-hand [5] according to the position of the laser with respect to the camera(s). An exception is [13] where a laser scan was rigidly attached to a camera. Thus, the authors considered the camera+laser system as degenerate stereoscopic system to control the depth directly by servoing the laser spot position on its epipolar line.

In all cases, the laser is considered as an additional sensor for the estimation of scene depth. Extending this idea, we propose to consider the moving laser as a virtual camera grabbing one pixel at a time. This allowed us to take into account the epipolar geometry constraint generated by the stereoscopic pair camera/laser in the design of more efficient visual servoing control laws [11]. Epipolar geometry-based visual servoing was efficiently investigated in the past [12], [14]–[16]. This type of control has also been successfully used for surgical applications [17], [18]. However, the more interesting idea for visual servoing applications is to use three views instead of two views. This is commonly known under the term trifocal system [19]. Some authors have addressed this problem in control approaches by coupling multiview geometry and visual servoing, for example using the trifocal constraint between two successive stereoscopic pairs [20] or between the initial, current and desired monocular views [21].

In this paper, the two-views and one laser system is considered as an analogous system to a degenerate case of trifocal geometry (Fig. 2). Thereby, it is possible to design original trifocal-based visual servoing approach simply using
the geometric properties of the three-view system. It will be shown that this approach has several advantages in terms of robustness, behavior, convergence or numerical complexity over the traditional approaches.

This paper is structured as follows: in Section II, trifocal geometry principles are provided and Section III deals with the proposed analogy of our system with the traditional trifocal system. Section IV shows design of the trifocal geometry based control law. Section V presents the experimental set-up and the validation results.

II. TRIFOCAL GEOMETRY

A. Notations

Let three cameras with optical centers $c_0$, $c_L$ and $c_R$ observe a 3D point $\mathbf{P} = (X, Y, Z)^T$ which is projected in 2D points $p = (x, y)^T$, $p_L = (x_L, y_L)^T$, and $p_R = (x_R, y_R)^T$ in the images planes $\Phi_0$, $\Phi_L$, and $\Phi_R$, respectively. The lines $(e_0 c_L)$, and $(e_0 c_R)$ define the baselines between each pair (camera, laser). The intersections of these baselines with the image planes give the epipole points $e_0$, $e_L$, and $e_R$. Also, $(e_0 p)$, $(e_L p_L)$, and $(e_R p_R)$ are called epipolar lines (Fig. 3). There are mathematical relations between the epipolar lines $(e_0 p)$, $(e_L p_L)$, and $(e_R p_R)$ and the 2D point $p$, commonly called epipolar constraint which are given by:

$$\mathbf{p}_L^T \mathbf{F}_L \mathbf{p}_L = 0 \quad (1)$$
$$\mathbf{p}_R^T \mathbf{F}_R \mathbf{p}_R = 0 \quad (2)$$
$$\mathbf{p}_L^T \mathbf{F}_R \mathbf{p}_R = 0 \quad (3)$$

where $\mathbf{F}_j$ is the fundamental matrix between camera $i$ and camera $j$ and $\mathbf{p}_i$ the homogenous representation of $p_i$ ($\mathbf{p}_i = (p_i, 1)^T$).

For a better understanding of different notations introduced in this paper, the reader may refer to the following Table 1.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_L$ and $I_R$</td>
<td>left image and right image, respectively</td>
</tr>
<tr>
<td>$\mathbf{p}_L$ and $\mathbf{p}_R$</td>
<td>2D homogeneous points projected in $I_L$ and $I_R$</td>
</tr>
<tr>
<td>$\mathbf{F}_j$</td>
<td>3D homogeneous point in the world frame $\mathbf{P}_j$</td>
</tr>
<tr>
<td>$\mathbf{F}_R$</td>
<td>projection of $\mathbf{P}$ in the image $I_R$</td>
</tr>
<tr>
<td>$\mathbf{p}^<em>$ and $\mathbf{p}^</em>$</td>
<td>desired image position and feed-forward term</td>
</tr>
<tr>
<td>$\mathbf{T}_j$</td>
<td>fundamental matrix between the cameras $i$ and $j$</td>
</tr>
<tr>
<td>$\mathbf{T}_i$</td>
<td>a 3x3 trifocal tensor</td>
</tr>
<tr>
<td>$z_0$</td>
<td>one of the three 3x3 matrix of $\mathbf{T}$</td>
</tr>
<tr>
<td>$\mathbf{r}_z$</td>
<td>$z$ direction of the laser beam from the mobile mirror towards the vocal folds</td>
</tr>
<tr>
<td>$\mathbf{r}_z^*$</td>
<td>desired direction of the laser beam</td>
</tr>
<tr>
<td>$\mathbf{d}$</td>
<td>time-derivation of $\mathbf{r}_z$</td>
</tr>
<tr>
<td>$\mathbf{Z}$</td>
<td>distance between the mobile mirror and the vocal folds</td>
</tr>
<tr>
<td>$\mathbf{h}_R$ and $\mathbf{h}_L$</td>
<td>depth information of the scene</td>
</tr>
<tr>
<td>$\mathbf{D}$</td>
<td>normal vector to the epipolar planes</td>
</tr>
<tr>
<td>$\mathbf{d}$</td>
<td>differential kinematic matrix of the mirror</td>
</tr>
<tr>
<td>$\omega$</td>
<td>rotation velocity of the mirror</td>
</tr>
</tbody>
</table>

TABLE I
NOTATIONS.

Fig. 1. Glottoscope in the past, today, and in the future.

Fig. 2. Schematic view of the laser steering system with two cameras.

Fig. 3. Trifocal geometry principle.
fined [19] by:

$$[\hat{\mathbf{p}}]_\times \left( \sum_{i=1}^{3} \hat{\mathbf{T}}_i \right) [\hat{\mathbf{p}}]_\times = 0_{3 \times 3}$$

(4)

where $\times$ is the cross-product and the set of $3 \times 3$ matrix $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_3$ represent the trifocal tensor $\mathbf{T}_{3 \times 3 \times 3}$. Therefore, it is possible to write the point-point-point transformation as:

$$[\hat{\mathbf{p}}]_\times (\hat{\mathbf{T}}_i^i) [\hat{\mathbf{p}}]_\times = 0_{3 \times 3}$$

(5)

with $i,j,k \in [1;3]$.

In the following section, by isolating the specific role of the virtual image, we introduce another way of writing these constraints, more appropriate for designing the trifocal geometry-based visual servoing.

### III. VIRTUAL TRIFOCAL GEOMETRY

As shown in Fig. 4, it is possible to define an analogy between our system and the conventional three-view system in Fig. 3. The approach shown here is based on a vectorial writing of the visual servoing problem using the trifocal constraint.

Let us note $\hat{0}_{\mathbf{z}}$ the direction (unit vector) of the laser beam/line from the actuated mirror towards the target (tissue), $\hat{\mathbf{P}}$ the 3D position of the laser spot on the tissue, and $\hat{\mathbf{p}}$ its projection in the image. They are related by:

$$\hat{\mathbf{P}} = d \hat{0}_{\mathbf{z}} = Z\hat{\mathbf{p}}$$

(6)

with $d$ the distance between the mobile mirror and the surface, $\hat{\mathbf{p}}$ is the projection of $\hat{\mathbf{P}}$ on the virtual image and $Z$ the depth (i.e., the distance parallel to the virtual optical axis). It has to be noticed that neither the distance $d$ nor the depth $Z$ are easy to determine unlike the vector $\hat{0}_{\mathbf{z}}$ that can obtained from the stereoscopic system, as shown later. We will not need $d$ and $Z$ any more in this paper.

From (6), it is easy to see that $\hat{0}_{\mathbf{z}}$ and $\hat{\mathbf{p}}$ are projectively the same. Therefore, it is possible to write:

$$\hat{0}_{\mathbf{z}} \equiv \hat{\mathbf{p}}$$

(7)

Actually, $\hat{\mathbf{p}}$ is the perspective projection on a plane, while $\hat{0}_{\mathbf{z}}$ is the perspective projection on a sphere.

### A. Trifocal Approach

An original visual servoing control law can be derived using the trifocal constraint offered by our multiview vision system. Similar to the epipolar constraint, the three-view geometry can also be used to exploit the geometrical aspects in designing the control law. Therefore, from the layout shown in Fig. 4, it is possible to rewrite the trifocal constraint (1), (2), (3) and (7) in the following form:

$$\begin{align*}
\hat{\mathbf{p}}_L^T \mathbf{F}_R \hat{\mathbf{p}}_R &= 0 \\
\hat{0}_{\mathbf{z}}^T \hat{\mathbf{p}}_R \hat{\mathbf{p}}_R &= 0 \\
\hat{0}_{\mathbf{z}}^T \mathbf{F}_L \hat{\mathbf{p}}_L &= 0
\end{align*}$$

(8)

(9)

(10)

Let us have aside (8) (except for improving the tracking of $\hat{\mathbf{p}}_L$ and $\hat{\mathbf{p}}_R$) and introduce $h_R$ and $h_L$ as:

$$\begin{align*}
h_R &= \hat{\mathbf{F}}_R \hat{\mathbf{p}}_R \\
h_L &= \hat{\mathbf{F}}_L \hat{\mathbf{p}}_L
\end{align*}$$

(11)

(12)

Consequently, from (9) and (10), it is trivial to see that $h_R \times h_L$ is parallel to $\hat{0}_{\mathbf{z}}$ (i.e. $\hat{0}_{\mathbf{z}} \equiv h_R \times h_L$). Using the “cross-product writing”, it is allowed to write:

$$\hat{0}_{\mathbf{z}} \times (h_R \times h_L) = 0$$

(13)

Note that this represents the trifocal constraint expressed in unbalanced manner, since $\hat{0}_{\mathbf{z}}$ is preferred.

### IV. VISUAL SERVOING BASED ON VIRTUAL TRIFOCAL GEOMETRY

#### A. Kinematic Transformations

The time derivative of (13) is:

$$\begin{align*}
\hat{0}_{\mathbf{z}} \times (h_R \times h_L) + \hat{0}_{\mathbf{z}} \times \left( \left( \mathbf{F}_R \hat{\mathbf{p}}_R \right) \times h_L \right) \\
+ \hat{0}_{\mathbf{z}} \times \left( h_R \times \left( \mathbf{F}_L \hat{\mathbf{p}}_L \right) \right) = 0
\end{align*}$$

(14)

which can be reorganized as:

$$\begin{align*}
\hat{0}_{\mathbf{z}} \times (h_R \times h_L) = \\
\hat{0}_{\mathbf{z}} \times \left( h_L \times (\mathbf{F}_R \hat{\mathbf{p}}_R) - h_R \times (\mathbf{F}_L \hat{\mathbf{p}}_L) \right)
\end{align*}$$

(15)

From the trifocal constraint (13), we also get trivially:

$$\hat{0}_{\mathbf{z}} = \frac{h_R \times h_L}{\| h_R \times h_L \|}$$

(16)

because $\hat{0}_{\mathbf{z}}$ is an unit vector. Hence, we have established a relationship between the velocities of the laser spot in each real image ($\hat{\mathbf{p}}_L$ and $\hat{\mathbf{p}}_R$) and the laser beam change of orientation $\hat{0}_{\mathbf{z}}$:

$$\begin{align*}
\hat{0}_{\mathbf{z}} \times \hat{0}_{\mathbf{z}} = \\
\frac{\hat{0}_{\mathbf{z}} \times (h_L \times (\mathbf{F}_R \hat{\mathbf{p}}_R) - h_R \times (\mathbf{F}_L \hat{\mathbf{p}}_L))}{\| h_R \times h_L \|}
\end{align*}$$

(17)
B. Proposed Control

Now, it is necessary to link the laser beam velocities $\frac{\partial \hat{z}}{\partial t}$ to those of laser spot denoted $\omega$ as follows:

$$0 = 2\omega \times \frac{\partial \hat{z}}{\partial t}$$  \hspace{1cm} (18)

where, the factor 2 comes from the reflection law in optics in the mirror.

The analytic solution to (18) is:

$$\omega = \frac{\partial \hat{z}}{\partial t} \times \frac{\partial \hat{z}}{\partial t} + k \frac{\partial \hat{z}}{\partial t}, \ k \in \mathbb{R}$$  \hspace{1cm} (19)

However, a rotation of the laser beam around its axis does not have any use, and so, $k$ is chosen to be 0.

Therefore, by identifying (17) and (19), we can easily deduce the expression of the new control law:

$$\omega = \frac{1}{2} \frac{\partial \hat{z}}{\partial t} \times \left( h_L \times \left( \hat{F}_R \hat{p}_R - h_R \times \left( \hat{F}_L \hat{p}_L \right) \right) \right)$$  \hspace{1cm} (20)

For an exponential decay of the cost-function, we introduce a first-order behavior of the error, in both images, between the current position $\hat{p}$ and the desired position $\hat{p}^*$ of the laser spot:

$$\dot{\hat{p}}_i = -\lambda (\hat{p}_i - \hat{p}^*_i) - \dot{\hat{p}}^*_i, \ i \in \{L,R\}$$  \hspace{1cm} (21)

where $\lambda$ is a positive gain and $\dot{\hat{p}}^*_i$ is the feed-forward term in case of trajectory tracking which is directly given by the surgeon through a smart human-machine interface [22].

Replacing (21) in (20) yields:

$$\omega = -\lambda \frac{\partial \hat{z}}{\partial t} \times \left( h_L \times \left( \hat{F}_R (\hat{p}_R - \hat{p}^*_R) \right) - h_R \times \left( \hat{F}_L (\hat{p}_L - \hat{p}^*_L) \right) \right)$$

$$\omega = -\kappa \times \left( h_L \times \hat{F}_R \hat{p}_R - h_R \times \hat{F}_L \hat{p}_L \right)$$  \hspace{1cm} (22)

where $\kappa = \frac{1}{||h_L||^2}.$

Finally, $\omega$ is converted to joint velocities $\dot{q}$ through the inverse differential kinematic matrix $D^{-1}$ of the mirror mechanism as $\omega = D^{-1} \dot{q}$.

C. Analysis

It can be observed that the design of this new controller requires only vector calculus. This involves using the multiview geometry specificities in another innovative and trivial manner. As shown by the expression of the new visual servoing control law (22), it does not involve any of the following: i) matrix inversion; ii) prior knowledge of the scene structure; iii) camera nor Euclidean hand-eye calibration; iv) explicit 3D reconstruction nor estimation of the scene; v) proprioceptive sensing to know the mirror configuration. All the data is directly obtained from the images. The only parameters required are the fundamental matrices $\hat{F}_R$ and $\hat{F}_L$. In the following section, it will be shown that using a coarse estimation of these matrices will be sufficient and allows for a more than satisfactory behavior (robustness, convergence, precision, etc.) of the proposed control law.

V. EXPERIMENTAL RESULTS

A. Experimental Set-up

The proposed approach was tested on an experimental set-up (Fig. 5) which consists of two CCD cameras characterized with a frame rate of 25 fps and a resolution of $480 \times 640$ pixels for each, a laser source, a fixed mirror, an actuated mirror (S-334) from Physical Instruments Inc. The latter contains two single axis units ($\alpha$ and $\beta$) working in series with one common pivot point characterized by a bandwidth of $200 \text{ Hz}$, a resolution of $0.2 \text{ \mu rad}$ and a motion range of $\pm 1.5^\circ$.

![Fig. 5. Photography of the experimental set-up.](image)

The current operational mode of the experimental set-up is as follows: the operator defines the desired position $\hat{p}^*_i$ (using a mouse click) in $I_L$, and using the fundamental matrix $\hat{F}_R$ between the left camera and the right camera, it is possible to find the corresponding desired position $\hat{p}^*_R$ in $I_R$. Then, the time-varying positions $\hat{p}_L$ and $\hat{p}_R$ of the laser spot in both images are tracked using ViSP [23].

B. Coarse Calibration

It is important to note again that the results described in this section are obtained without any camera nor eye-to-hand calibration (i.e., left camera/mirror and right camera/mirror poses). The fundamental matrices $\hat{F}_L$ (between the laser beam frame $R_0$ linked to the reflexion point in the mobile mirror and the left camera frame $R_L$), and $\hat{F}_R$ (between $R_0$ and the right camera frame $R_R$) are approximated (estimated at hand) by a translation vector $\hat{\omega}_{L,R} = (1, -0.1, 38)^\top \text{ mm}$ and a rotation matrix $\hat{R}_{L,R} = \mathbf{0}_{3 \times 3}$ and $D$ identified in the image as $D = \begin{bmatrix} -10 & 0 & 0 \\ 0 & 10 & 0 \end{bmatrix}$.

C. Results

Not being able yet to test our developments directly on true vocal cords due to the lack of a technical solution for fiber imaging, we have imagined two scenarios. First, the surgical site consists of a planar scene on which the laser spot is projected. Second, the 2D scene is replaced by a 3D shape phantom.
Fig. 6. Some images captured by the left camera which illustrate the displacement of the laser spot during the trifocal visual control. Superimposed lines are the performed trajectories and the ideal one.

Fig. 7. The same task shown in Fig. 6 viewed by the right camera.

Fig. 8. (a) Image errors evolution vs. iterations $i$ and (b) velocities evolution vs. iterations $i$ [$I_L$].

Fig. 9. Image sequence captured by the left camera during the laser steering on a 3D shape phantom. Superimposed lines are the performed trajectories and the point-to-point straight line.

The laser steering is also validated in a scenario closer to the final application in order to study the performance of the controller in such conditions of use. To do this, we have manufactured a 3D shape surface (using a 3D printer) with hills of 5 mm of height. However, the validation task remains the same as in the 2D surface. Obviously, the idea is to observe the behavior of the laser spot displacements in this 3D object under the same conditions as in the first scenario i.e., the same set-up configuration and the coarse calibration parameters.

The results of this second scenario are shown in Fig. 9 and 10 which represent some images captured by the left and the right cameras, respectively. They show the same task performed on a 3D surface. It can be seen that the laser spot point-to-point trajectories are very close to a straight-line. This is performed without any prior knowledge nor 3D estimation (e.g., 3D reconstruction or depth computation) on the used 3D surface.

As in the case of 2D scene experiments, we can also notice the fact that the controller presents near perfect exponential convergence and decoupling behaviors. This is observed
camera which will be connected to two image bundles through miniature GRIN lenses. It will be concern the study of the accuracy of the control law (using external sensor) and its stability (global proof of stability).

REFERENCES