

RESEARCH ARTICLE

Visual servoing of a laser beam through a mirror

Japheth Ka'pesho Odira^{1,2,*} , Nicolas Andreff¹, Luc Petiet¹ and Jean Bosco Byiringiro²

¹Automatic Control and Micro-Mechatronic Systems Department (AS2M), FEMTO-ST Institute, CNRS-UFC/ENSMM, Besancon 25000, France and ²Siemens Mechatronics Certification Centre, Dedan Kimathi University of Technology, Nyeri, Kenya

*Corresponding author. E-mail: japhethkapesha@gmail.com

Received: 30 April 2021; **Revised:** 6 December 2021; **Accepted:** 12 January 2022; **First published online:** 24 March 2022

Keywords: minimally invasive surgery, vocal fold access, visual servoing, mirror, laser beam steering

Abstract

In this study, we present a new approach to improving vocal fold access to perform phonomicrosurgery. It is done by shooting the laser through a mirror to reach the vocal fold hidden parts. A geometrical study of laser shooting path was conducted for vocal fold anatomical constraints, followed by devising a laser-shooting system conceptual design. Control laws were developed and tested by simulation and validated experimentally on a test bench in a monocular and stereoscopic configuration. Simulation and experimental results are provided to demonstrate the effectiveness of the developed approach.

1. Introduction

The demand to improve health quality has led to much research including, phonomicrosurgeries, which involves delicate surgical operations on the vocal fold and requires a highly skilled surgeon [1, 2]. Vocal fold surgery requires precision and accuracy due to the tissue's nature being resected, thin, fragile, and viscous. Their lesions might be less than 1 mm [3, 4]. The most common procedure to resect those lesions relies on laser surgery.

Systems for laser phonosurgery, such as Acupulse Duo by Lumenis [5], are based on a laser manipulator mounted onto an external microscope. The patient is placed in extreme neck extension so that a rigid straight laryngoscope can be placed in the patient mouth and throat to allow a direct line of sight between the laser manipulator and vocal fold in the larynx. However, certain vocal fold portions are inaccessible in such a placement as lateral and posterior sides because the laser source is located out of the patient's body. A small area laser beam could be moved into the laryngoscope, which prevents the surgeon from conducting a surgical operation on those portions.

Another system, a flexible endoluminal robotic system, was developed during the European project μ RALP [6]. That concept has a miniaturized laser manipulator, having micro cameras embedded at the endoscope tip and shooting laser from within the larynx. Nevertheless, since the laser source was placed above the vocal fold and light only travels along straight lines hindered the surgeon from operating on the anterior vocal fold but hardly had access to lateral sides and, to the best of our knowledge, no access to the posterior side. The μ RALP project also proposed improving laser steering accuracy by automatically controlling the laser [7] to follow a surgeon-drawn path [8], rather than having the surgeon manually steer the laser beam through a poorly ergonomic joystick. This automatic control is done by visually servoing laser spots from one [7] or two [9, 10] endoscopic cameras.

There are several works reported in the literature, especially in visually guided laser ablation catheter [11], which was designed to allow the operator to directly visualize target tissue for ablation and then deliver laser energy to perform point-to-point circumferential ablation. Also, the velocity independent visual path following for laser surgery in ref. [12] where nonholonomic control of the unicycle

model and path following at high frequency to satisfy the constraints of laser–tissue interaction was explored. Another example is reported in ref. [13], where a robotic system for skin photo-rejuvenation, which uniformly delivers laser thermal stimulation to a subject's facial skin tissue, was investigated. Yet, as far as we could understand it, none of those work automatically steered laser along hidden paths.

Visual servoing techniques use visual information extracted from the images to design a control law [14–16]. It is a systematic way to control a robot using the information provided by one or multiple cameras [17]. Standard stereo sensors used for visual servoing have a limited view and consequently limit their application range. Hence, using planar mirrors has been prioritized to enlarge the field of view of classic pinhole cameras [18] and for high-speed gaze control [19, 20]. A planar mirror is a mirror with a planar reflective surface. An image of an object in front of it appears to be behind the mirror plane. This image is equivalent to one being captured by a virtual camera located behind the mirror; additionally, the virtual camera's position is symmetric to the position of the real camera. In our case, the camera's reflection in the planar mirror is used to track virtual features points in the vocal fold hidden scene. For instance, by using mirror reflections of a scene, stereo images are captured [21].

Tracking is the problem of estimating the trajectory of an object in the image plane as it moves around a scene. A tracker allocates unswerving labels to the tracked objects in different video frames. In addition, depending on the tracking field, a tracker can provide object-centric information, such as the area or shape of an object. Simple algorithms for video tracking rely on selecting the region of interest in the first frame associated with the moving objects. Tracking algorithms can be classified into three categories: point tracking [22, 23], kernel tracking [24, 25], and silhouette tracking [26]. Occlusion can significantly undermine the performance of the object tracking algorithms. Occlusion often occurs when two or more objects come too close and seemingly merge or combine. Image processing systems with object tracking often wrongly track the occluded objects [27]. After occlusion, the system will wrongly identify the initially tracked objects as new objects [28]. If the geometry and placement of static objects in the real surroundings are known, the so-called phantom model is a common approach for handling the occlusion of virtual objects. A method for detecting a dynamic occlusion in front of static backgrounds is described in ref. [29]. This algorithm does not require any previous knowledge about the occluding objects but relies on a textured graphical model of planar elements in the scene. Some approaches solve the occlusion problem using depth information delivered by stereo matching [30]. In our approach to the occlusion problem, we use the triangulation method, where we pay attention to the pixels which are well reconstructed when an image is reproduced and ignore the ones which are not well reconstructed.

The study focussed on a conceptual method of servoing laser to hidden parts of the vocal fold. Having been inspired by this clinical need for improved access and those work pieces on mirror reflections, we propose an analysis of anatomical constraints of the vocal fold, devised a conceptual design, and formulated a controller, which was evaluated experimentally on a tabletop setup.

The first contribution is to propose a method to access parts of the vocal fold workspace that are not directly visible during phonomicrosurgery, for instance, the posterior side of the vocal fold by seeing through an auxiliary mirror to overpass the limited micro-cameras field of view of a flexible endoscopy system that was missing in ref. [7], and shooting surgical laser using the same auxiliary mirror to access those invisible parts in the vocal fold workspace. The second contribution is to derive the control equations for automatically steering the laser through the auxiliary mirror to the hidden parts of the vocal fold by updating the control in refs. [9, 10]. However, through modelling, simulation, and experimentation, the addition of the auxiliary mirror is shown to have no impact, which is thereby demonstrated as being used as it is. Nonetheless, we took the opportunity of this study to derive a variant of the control in ref. [9] based on the geodesic error (cross-product) rather than on the linear error. Fig. 1 shows a sample of a simulated image of the vocal fold.

The remainder of this study is presented as follows. Section 2 gives a detailed description of the conceptual system design to access parts of the vocal fold. Section 3 deals with modelling the proposed

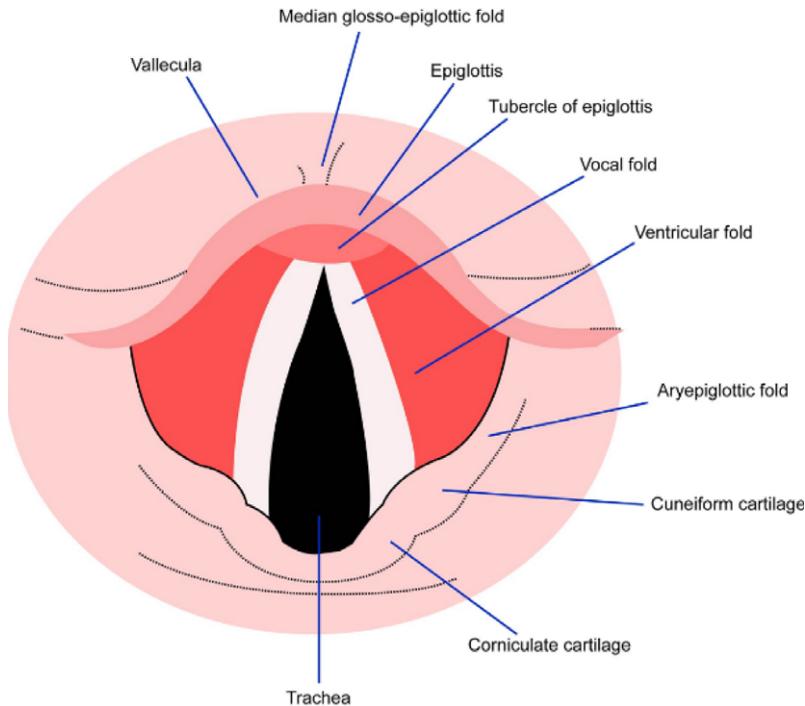


Figure 1. Simulated image of vocal fold anatomy.

system both in monocular and stereoscopic cases to establish a controller. Section 4 focuses on the simulation results of the controller for both the monocular case and stereoscopic case. Section 5 presents the performed experimental validations in a tabletop setup.

2. Design

2.1. System configuration

As illustrated in Fig. 2, the objective is to devise a method that improves access to hidden parts of the vocal fold. In our approach, we propose a system with two cameras to give stereo view and visual feedback to the scene, a laser source to provide surgical laser needed for tissue ablation, illumination guidance, auxiliary mirror guide, and an auxiliary mirror manipulator through which laser is steered to hidden parts of the vocal fold. In practice, human tissues will never contact the designed micro-robotic device, just the endoscope outer shell, which can be readily sterilized and biocompatible. An auxiliary mirror would be inserted at the beginning of the surgical process and remain stationary until the end of surgery. All those parts must be miniaturized and enclosed in a flexible endoscope during fabrication which is out of this paper's scope. However, details on packing all those hardware components (miniaturized) into an endoscope can be found in ref. [31]

2.2. System model for accessing hidden parts of vocal fold

From the system configuration above, enlarging its distal arrangements of a flexible endoscope and focussing on how to access hidden features of the vocal fold is shown in Fig. 3. The system has a micro-robot, a tip/tilt actuating mirror to steer the laser through an auxiliary mirror to reach hidden parts. The two cameras also observe the same hidden scene through a mirror reflection. Hence, providing a clear vision of the surgeon's stage to define a trajectory followed automatically by surgical laser in those hidden parts.

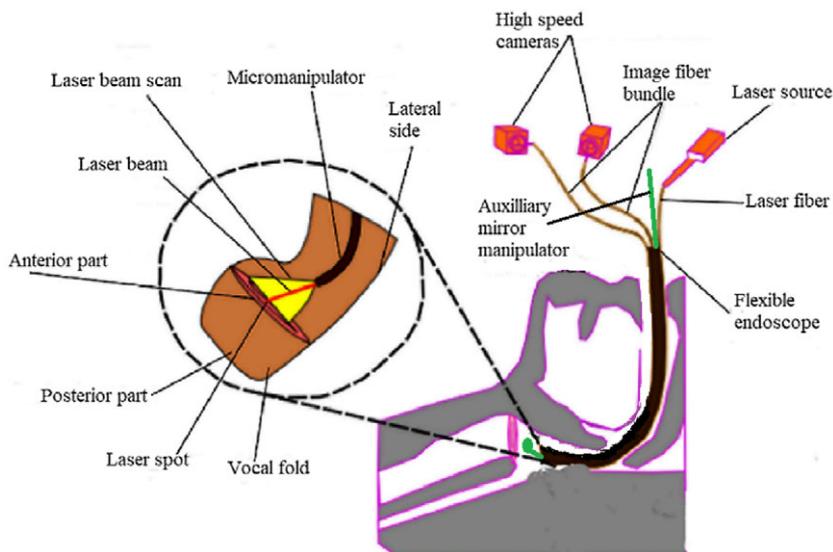


Figure 2. System configuration.

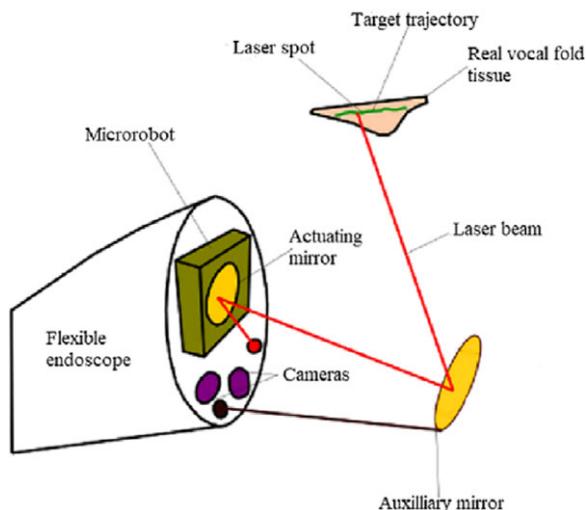


Figure 3. System model.

2.3. Enlarged anterior, lateral, and posterior view for accessing hidden parts of vocal fold

The former configuration of Fig. 2 has a limited field of view, even on anterior parts of the vocal fold, because of technical constraints (miniaturization for endoluminal systems, direct line of sight for extracorporeal systems). Thus, the proposed method for accessing hidden parts on the vocal fold anterior side is in Fig. 4(a). Based on the orientation and position of the auxiliary mirror, the surgical laser can be steered to all parts on the anterior side of the vocal fold, which is the surface of the vocal fold that is visible from the larynx. For instance, the laser can reach parts outside of the direct view field (in yellow).

In Fig. 4(b), a surgical laser is first shot to an auxiliary mirror, then reflected towards the tissues located above the vocal fold in the larynx, such as the ventricular fold.

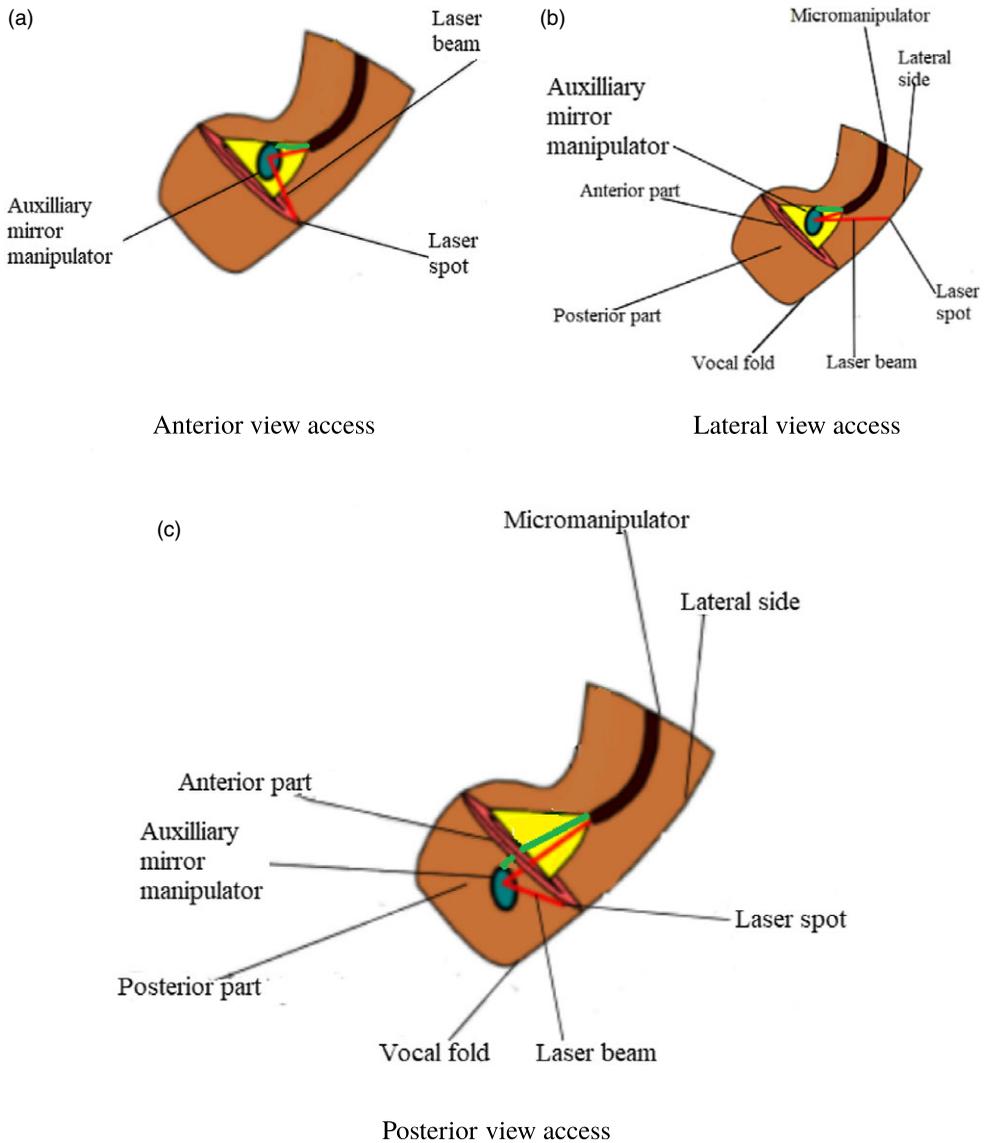


Figure 4. (a) Anterior view access. (b) Lateral view access. (c) Posterior view access.

As demonstrated in Fig. 4(c), rare portions of a vocal fold being the surface visible from the trachea can be accessed by opening the vocal lips using forceps, exposing the backside for the auxiliary mirror to be oriented and pushed.

3. Modelling

3.1. Mirror reflection

From a technical point of view, these control equations also differ from the already published ones [7, 9, 10] by the use of an alternative formulation of the perspective projection model and by the servoing of geodesic image errors instead of linear image errors. Table I shows list of symbols used in the paper.

Table I. List of symbols used in the paper.

Symbol	Remarks
π	Mirror plane
\underline{n}	Vector normal to the mirror plane
d	Distance of the reference frame origin to the mirror plane
\tilde{X}_R	Homogeneous coordinates for the reflected point
\tilde{X}	Homogeneous coordinates for the real point
\mathbf{D}	Reflection matrix
${}^w\tilde{X}_R$	Homogeneous coordinates for the reflected point in the world frame
${}^c\tilde{X}_R$	Homogeneous coordinates for the reflected point in the camera frame
$\mathbf{I}_{3 \times 4}$	Canonical perspective projection in the form of 3×4 identity matrix
cT_w	Euclidean 3D transformation
cG_w	Notation simplification of the product of identity, transformation matrix, and reflection matrix
${}^c\tilde{x}$	A 2D point position in the image coordinate
$\dot{{}^c\tilde{x}}$	Velocity of the 2D point in the image coordinate
${}^w\dot{\tilde{X}}$	Velocity of the 3D point
δ_c	The unknown depth along the line of sight passing through ${}^c\tilde{x}$
mz	The virtual projected spot on the mirror virtual image plane
mT_w	Transformation matrix relating micro-mirror frame (m) with world frame through the auxiliary mirror
${}^m\dot{z}$	The velocity of the virtual projected spot on the mirror virtual image plane
e_{geo}	The geodesic error
${}^c\tilde{x}$	The detected position of the laser spot in the image
${}^c\tilde{x}^*$	The desired position of the laser spot in the image
ζ	a pseudo-control signal on the sphere
λ	a positive gain
ω	Angular velocity of the actuated mirror
${}^o\tilde{p}$	Homogeneous representation of 2D point
oF_L	Left fundamental matrix
oF_R	Right fundamental matrix
h_R and h_L	Epipolar lines in right and left images, respectively

Consider the reflection of X into X_R through the auxiliary mirror plane $\pi = (\underline{n}^T, d)^T$ where \underline{n} is the vector normal to the mirror plane, and d is the distance of the reference frame origin to the mirror plane. Using homogeneous coordinates for the points and following [32], one has

$$\tilde{X}_R = \mathbf{D}\tilde{X} \tag{1}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} - 2\underline{n}\underline{n}^T & 2d\underline{n} \\ \mathbf{0}_{3 \times 1}^T & 1 \end{bmatrix} \tag{2}$$

Implementing these equations depends on the chosen reference frame and can thus be expressed either in the world frame.

$${}^w\tilde{X}_R = {}^w\mathbf{D}{}^w\tilde{X} \tag{3}$$

or in a camera frame

$${}^c\tilde{X}_R = {}^c\mathbf{D}{}^c\tilde{X} \tag{4}$$

3.2. Camera projection based on a cross-product concept

When a camera captures an image of a scene, depth information is lost as objects, and points in 3D space are mapped onto a 2D image plane. For the work in this study, depth information is crucial since there is a need for scene reconstruction from the information provided by the 2D image to know the distance between the actuated mirror and the scene without prior knowledge of where they are. Therefore, the used approach of the perspective (pinhole) image projection \tilde{x} of a 3D point \tilde{X} is stated as

$${}^{im}\tilde{x} \equiv {}^{im}\tilde{\mathbf{K}}_c \mathbf{I}_{3 \times 4} {}^c\mathbf{T}_w {}^w\mathbf{D} {}^w\tilde{X} \tag{5}$$

where ${}^{im}\tilde{\mathbf{K}}_c$ represents calibrated intrinsic camera parameters, $\mathbf{I}_{3 \times 4}$ represents a canonical perspective projection in the form of 3×4 identity matrix, ${}^c\mathbf{T}_w$ represents Euclidean 3D transformation (rotation and translation) between the two coordinate systems of camera and world through a mirror. The \equiv sign represents depth loss in the projection up to some scale factor. In practice, the \equiv sign can be removed through division operation, which introduces non-linearity. Alternately, using the cross product, we can make the projection equation a linear constraint equation. Since the light ray emitted from the camera centre point aligns with the light ray coming from the 3D point. If we treat light ray emitted from the camera centre point as vector \mathbf{A} and light ray coming from the 3D point as vector \mathbf{B} . The cross-product of two 3D vectors \mathbf{A} and \mathbf{B} gives another vector with a magnitude equal to that of the area enclosed by the parallelogram formed between the two vectors. The direction of this vector is perpendicular to the plane enclosed by \mathbf{A} and \mathbf{B} in the direction given by the right-hand rule, and the magnitude of the cross product will be given by $|\mathbf{A} \times \mathbf{B}| \sin \theta$. However, if these two vectors are in the same direction, just like in our case, the angle between them will be zero. The magnitude of the cross product will be zero since $\sin(0) = 0$. The resultant vector will be the zero vector. $\mathbf{A} \times \mathbf{B} = 0$

$$[{}^{im}\tilde{\mathbf{K}}_c^{-1} {}^{im}\tilde{x}]_{\times} \mathbf{I}_{3 \times 4} {}^c\mathbf{T}_w {}^w\mathbf{D} {}^w\tilde{X} = 0 \tag{6}$$

Hence for notation simplicity, let

$${}^c\tilde{x} = {}^{im}\tilde{\mathbf{K}}_c^{-1} {}^{im}\tilde{x} \tag{7}$$

$${}^c\mathbf{G}_w = \mathbf{I}_{3 \times 4} {}^c\mathbf{T}_w {}^w\mathbf{D} \tag{8}$$

Consequently, Eq. (6) rewrites more simply as

$$[{}^c\tilde{x}]_{\times} {}^c\mathbf{G}_w {}^w\tilde{X} = 0 \tag{9}$$

3.3. Laser spot kinematics

The time derivative of Eq. (9) is considered to servo the spot position from the current position to the desired one while the camera and mirror remain stationary.

$$[{}^c\dot{\tilde{x}}]_{\times} {}^c\mathbf{G}_w {}^w\tilde{X} + [{}^c\tilde{x}]_{\times} {}^c\mathbf{G}_w {}^w\dot{\tilde{X}} = 0 \tag{10}$$

Since ${}^c\tilde{x}$ is a unit vector (i.e. $\|{}^c\tilde{x}\| = 1$) and using Eq. (9) yields

$${}^c\mathbf{G}_w {}^w\tilde{X} = \delta_c {}^c\tilde{x} \tag{11}$$

with $\delta_c > 0$ the unknown depth along the line of sight passing through ${}^c\tilde{x}$ hence, Eq. (10) becomes

$$[{}^c\tilde{x}]_{\times} {}^c\dot{\tilde{x}} = \frac{1}{\delta_c} [{}^c\tilde{x}]_{\times} {}^c\mathbf{G}_w {}^w\dot{\tilde{X}} \tag{12}$$

3.4. Scanning laser mirror as a virtual camera

Scanning a mirror as a virtual camera is considered with a virtual image plane. Therefore, the mathematical relationship between it and the 3D spot on the reflected vocal fold is established in Eq. (13)

below. Some parameters of Eq. (6) have changed as $c = m$, and $\mathbf{K} = \mathbf{I}_{3 \times 3}$ since when using the mirror as a camera, focal length, optical centre, and lens distortion are no longer a problem hence \mathbf{K} is taken to be one.

$$[{}^m z]_{\times} \mathbf{I}_{3 \times 4} {}^m \mathbf{T}_w {}^w \mathbf{D}^w \tilde{\mathbf{X}} = \mathbf{0} \tag{13}$$

${}^m z$ is the virtual projected spot on the mirror virtual image plane, ${}^m \mathbf{T}_w$ transformation matrix relating micro-mirror frame (m) with world frame through the auxiliary mirror (w) hence ${}^m \mathbf{T}_w$ constant. Differentiating Eq. (13) gives the velocity at which the laser is servoed from one point to another in the image. The resultant equation is

$${}^m \dot{z} = -\frac{1}{\delta_m} \mathbf{I}_{3 \times 4} {}^m \mathbf{T}_w {}^w \mathbf{D}^w \dot{\tilde{\mathbf{X}}} \tag{14}$$

3.5. To be virtual or not to be?

The overall static model for both the laser steering system through the auxiliary mirror and a camera observing the laser spot through *the same mirror* is given by the constraints in Eqs. (6) and (13). This forms an *implicit* model of the geometry at play, from which one can, depending on what is known beforehand and what's needed, *explicitly* try to get the unknown values from the known ones. The easiest is to find the laser direction and its spot projection in the image from a known place of the spot in 3D and the 3D locations of the camera ${}^c \mathbf{T}_w$, the steering mirror ${}^m \mathbf{T}_w$, and auxiliary mirror ${}^w \mathbf{D}$. However, in practice, one would like to “triangulate through the mirror” the 3D spot from the laser orientation and the spot image projection. And even more helpful, one would like to steer the laser (i.e., change z , thus \mathbf{X}) from an image-based controller (i.e., a desired motion of \mathbf{x}). Then, the question is whether one should explicitly reconstruct \mathbf{X} or can the controller be derived without this explicit reconstruction.

A large part of the answer to that question lies in the auxiliary mirror location ${}^w \mathbf{D}$. If it is known, then triangulation can potentially be done, but this imposes strong practical constraints. However, looking closely at the above equations and Fig. 4, one can remark that there exists a virtual spot location, $\tilde{\mathbf{X}}_R = {}^w \mathbf{D} \mathbf{x}$ which lies behind the mirror. Replacing ${}^w \mathbf{D} \mathbf{x}$ by $\tilde{\mathbf{X}}_R$ in Eqs. (6) and (13) yields a solution independent of the auxiliary mirror location.

$$[{}^c \tilde{\mathbf{x}}]_{\times} \mathbf{I}_{3 \times 4} {}^c \mathbf{T}_w {}^w \tilde{\mathbf{X}}_R = \mathbf{0} \tag{15}$$

$$[{}^m \tilde{z}]_{\times} \mathbf{I}_{3 \times 4} {}^m \mathbf{T}_w {}^w \tilde{\mathbf{X}}_R = \mathbf{0} \tag{16}$$

Of course, this simplification is only valid when *both the laser and the camera reflect through the same mirror*, forcing the user to check that the laser spot is visible in the image. This also reduces the calibration burden to determine the relative location ${}^c \mathbf{T}_m$ between the steering mirror and the camera since the steering mirror frame can arbitrarily be chosen as the world frame of the virtual scene. As a consequence, from a modelling point of view, working with the virtual scene reduces the problem to its core.

$$[{}^c \tilde{\mathbf{x}}]_{\times} \mathbf{I}_{3 \times 4} {}^c \mathbf{T}_m {}^m \tilde{\mathbf{X}}_R = \mathbf{0} \tag{17}$$

$$[{}^m \tilde{z}]_{\times} \mathbf{I}_{3 \times 4} {}^m \tilde{\mathbf{X}}_R = \mathbf{0} \tag{18}$$

As will be seen in the following sections, this allows to derive a controller without making an explicit triangulation, that is, without necessarily having sensors for z .

Consequently, placing the problem in the virtual space allows for a simple solution, independent from prior knowledge of the auxiliary mirror location, which just needs to be held stable during control so that the desired visual feature and the current one are geometrically consistent.

3.6. Geodesic error

Geodesic error differs from linear error since error reduction is made along the unit sphere’s surface for geodesic error rather than within the image plane, resulting in linear error minimization.

$$e_{geo} = {}^c\tilde{\mathbf{x}} \times {}^c\tilde{\mathbf{x}}^* \tag{19}$$

where ${}^c\tilde{\mathbf{x}}$ is the detected position of the laser spot in the image and ${}^c\tilde{\mathbf{x}}^*$ is the desired one, which is chosen arbitrarily by users in the visual image, and e_{geo} representing the shortest arc between the two points defining the rotation vector orthogonal to the arc plane. Once ${}^c\tilde{\mathbf{x}}$ is a unit vector, its derivative takes the form of $\dot{{}^c\tilde{\mathbf{x}}} = \zeta \times {}^c\tilde{\mathbf{x}}$ where ζ is a pseudo-control signal on the sphere and replacing in Eq. (12) yields

$$\zeta = \lambda e_{geo} = \lambda ({}^c\tilde{\mathbf{x}} \times {}^c\tilde{\mathbf{x}}^*) \quad \lambda > \mathbf{0} \tag{20}$$

$$\lambda [{}^c\tilde{\mathbf{x}}]_{\times}^3 {}^c\tilde{\mathbf{x}}^* = -\frac{1}{\delta_c} [{}^c\tilde{\mathbf{x}}]_{\times} {}^c\mathbf{G}_w {}^w\dot{\mathbf{X}} \tag{21}$$

since $[{}^c\tilde{\mathbf{x}}]_{\times}^3 = -[{}^c\tilde{\mathbf{x}}]_{\times}$ and the virtual 3D laser spot velocity ${}^m\dot{\mathbf{X}}_R$, to be controlled, is thus constrained by

$$\lambda e_{geo} = -\frac{1}{\delta_c} [{}^c\tilde{\mathbf{x}}]_{\times} {}^c\mathbf{G}_m {}^m\dot{\mathbf{X}}_R \tag{22}$$

3.7. Single-camera case of observing hidden portions of vocal fold

We can effectively model and control the laser path with one camera, actuating mirror, and auxiliary mirror. By first establishing angular velocity of the actuating mirror to control the orientation of the laser beam, the general solution to Eq. (22) is

$$\lambda \delta_c e_{geo} + \mathbf{k} {}^c\tilde{\mathbf{x}} = {}^c\mathbf{G}_m {}^m\dot{\mathbf{X}}_R \quad \mathbf{k} \in \mathbf{R} > \mathbf{0} \tag{23}$$

where $\mathbf{k} {}^c\tilde{\mathbf{x}}$ can be interpreted as the motion of \mathbf{X} along its line of sight (thus a variation of δ_c) that is not observable by the camera. It can be due to the irregular shape of the surface hit by the laser or made by a specific motion of that surface.

Observing that

$${}^c\mathbf{G}_m {}^m\dot{\mathbf{X}}_R = \mathbf{I}_{3 \times 4} \begin{bmatrix} {}^c\mathbf{R}_m & {}^c\mathbf{t}_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} {}^m\dot{\mathbf{X}}_R \\ 0 \end{bmatrix} = {}^c\mathbf{R}_m {}^m\dot{\mathbf{X}}_R \tag{24}$$

allows solving for ${}^m\dot{\mathbf{X}}_R$ in Eq. (23)

$${}^m\dot{\mathbf{X}}_R = {}^c\mathbf{R}_m^T (\lambda \delta_c e_{geo} + \mathbf{k} {}^c\tilde{\mathbf{x}}) \tag{25}$$

Hence, substituting ${}^m\dot{\mathbf{X}}_R$ with Eq. (25) result in

$${}^m\dot{\mathbf{z}} = -\frac{1}{\delta_m} {}^c\mathbf{R}_m {}^c\mathbf{R}_m^T (\lambda \delta_c e_{geo} + \mathbf{k} {}^c\tilde{\mathbf{x}}) \tag{26}$$

which simplifies into

$${}^m\dot{\mathbf{z}} = -\lambda' e_{geo} + \mathbf{k}' {}^c\tilde{\mathbf{x}} \tag{27}$$

where $\lambda' = \frac{\lambda \delta_c}{\delta_m}$ and $\mathbf{k}' = \frac{\mathbf{k}}{\delta_m}$ are the control gains and can be tuned without explicit reconstruction of the depths δ_c and δ_m . Again, the controller is independent of the mirror’s position because both the image and the laser go through it. \mathbf{k} can be taken as zero unless one wishes to estimate and compensate for the surface shape and ego motion.

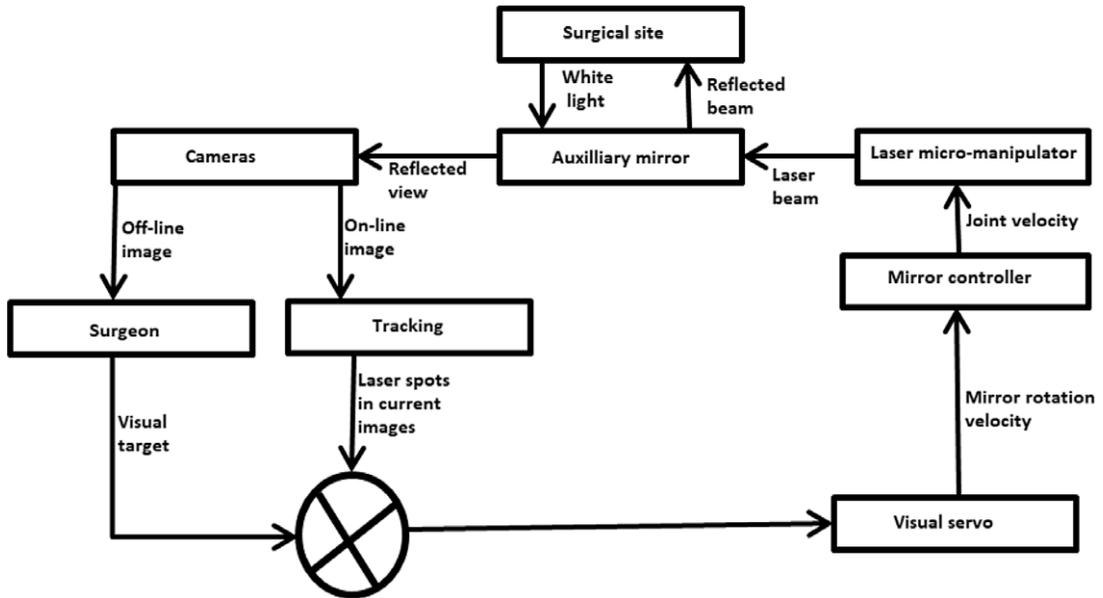


Figure 5. The system model workflow.

The relationship between laser speed velocity and angular velocity of the actuated mirror is given as [9]

$${}^o\dot{z} = \omega \times {}^o z \tag{28}$$

$$\omega \times {}^o z = -\lambda' e_{geo} \tag{29}$$

Making ω the subject of the formula from Eq. (29)

$$\omega = -\lambda' e_{geo} \times {}^o z \tag{30}$$

Fig. 5 shows the system model workflow.

3.8. Trifocal geometry

Let us now investigate the effect of using two cameras, in addition to the actuating mirror and an auxiliary mirror.

In Fig. 6, three cameras with optical centres c_o , c_L , and c_R observe a 3D point $P = (x, y, z)^T$ through a mirror as point P_{rf} which is projected in 2D points ${}^o p = (x, y)^T$, $p_L = ({}^L\tilde{x}, {}^L\tilde{y})^T$ and $p_R = ({}^R\tilde{x}, {}^R\tilde{y})^T$ in the images planes ϕ_o , ϕ_L and ϕ_R , respectively.

The fundamental matrices ${}^o F_R$ and ${}^o F_L$ and the epipolar lines e_L and e_R showing a relation between the cameras and actuated mirrors.

There are mathematical relations between the epipolar lines ($e_L p_L$) and ($e_R p_R$) and 2D point ${}^o p$, commonly called Epipolar constraints, which are given by

$${}^o\tilde{p}^{To} F_L^c \tilde{p}_L = 0 \tag{31}$$

$${}^o\tilde{p}^{To} F_R^c \tilde{p}_R = 0 \tag{32}$$

$${}^c\tilde{p}_L^{TL} F_R^c \tilde{p}_L = 0 \tag{33}$$

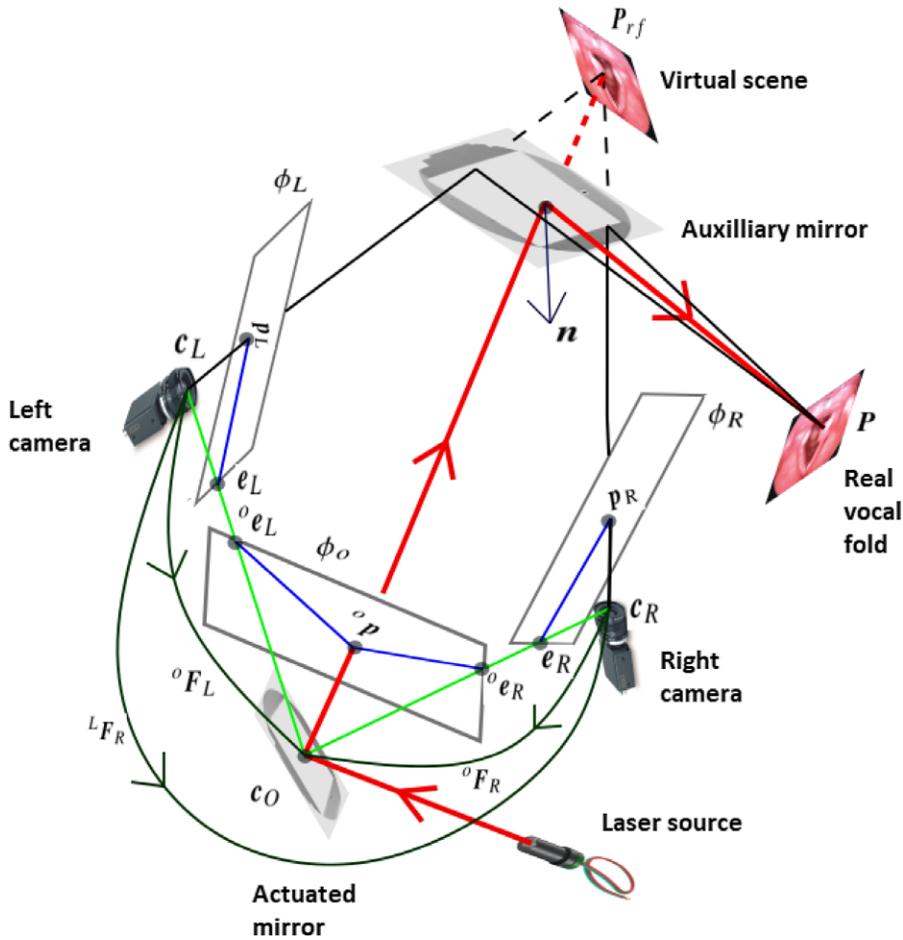


Figure 6. Model schematic.

Hence, mirror velocity as a function of the laser spot velocities is expressed as [9]

$$\omega = -\frac{\mathbf{h}_R \times \mathbf{h}_L}{\|\mathbf{h}_R \times \mathbf{h}_L\|^2} \times \left\{ \mathbf{h}_L \times \left({}^o\mathbf{F}_R^R \dot{\tilde{\mathbf{x}}} \right) - \mathbf{h}_R \times \left({}^o\mathbf{F}_L^L \dot{\tilde{\mathbf{x}}} \right) \right\} \tag{34}$$

where $\mathbf{h}_R = \mathbf{e}_R \times {}^o\mathbf{z}$ and $\mathbf{h}_L = \mathbf{e}_L \times {}^o\mathbf{z}$. However, contrary to ref. [9], where a linear error was used in both images, we use here geodesic error which involves error minimization along the surface of the object rather than error minimization between two points on the object.

$${}^c\mathbf{e}_{geo} = {}^c\tilde{\mathbf{x}} \times {}^c\tilde{\mathbf{x}}^* \quad c \in \{L, R\} \tag{35}$$

Its time derivative was as follows:

$${}^c\dot{\mathbf{e}}_{geo} = {}^c\dot{\tilde{\mathbf{x}}} \times {}^c\tilde{\mathbf{x}}^* \tag{36}$$

under the assumption that the desired configuration is piecewise constant. Then λ is a positive gain, and ${}^c\mathbf{e}_{geo}$ undergoes exponential decay to increase the convergence rate.

$${}^c\dot{\mathbf{e}}_{geo} = -\lambda {}^c\mathbf{e}_{geo} \tag{37}$$

Table II. *Simulation parameters.*

Fields	Quantity
Simulation period	2 milliseconds
Camera separation distance	5 mm
Vocal fold offset distance	28 mm
Vocal fold orientation	30 degrees (+)
Auxiliary mirror offset distance	25 mm
Auxiliary mirror orientation	45 degrees(-)

Hence substituting Eq. (37) into Eq. (36) yields

$${}^c\dot{\tilde{\mathbf{x}}} = -\lambda {}^c\tilde{\mathbf{x}} \times {}^c\mathbf{e}_{geo} + \mu {}^c\tilde{\mathbf{x}}^* \quad (38)$$

where μ can be chosen in different manners;

Here, we chose $\mu = 0$ and relied on the physical constraints; then, the final controller is obtained by applying Eq. (38) for each image and reporting the result into Eq. (34).

4. Simulation

4.1. Simulation results for a single camera and auxiliary mirror in a realistic case

ViSP was used for simulation in C++ SDK, and graphs were plotted with Octave. Table II shows parameters used during the simulation period.

As actual geometry of the vocal fold is not needed in the controller, since it is implicitly included in the desired and current laser spot position, we simplified the simulation of the vocal fold to a simple planar patch.

Figure 7 shows the simulation setup used. Where point S corresponds to the laser spot position on the vocal fold. Vector \mathbf{u} is a unit vector in the direction of a laser beam, \mathbf{d} is the shortest distance from the centre of the micro-mirror to the plane of the auxiliary mirror. Vector \mathbf{u}_2 is the reflected unit vector of \mathbf{u} , \mathbf{d}_1 is the shortest distance between the auxiliary mirror and vocal fold plane, and \mathbf{z}_2 is the distance along the reflected laser beam.

This simulation aims to validate the laser monocular visual servoing through an auxiliary mirror, controlled by Eq. (30). Figure 8(a) orange colour asterisk is the laser spot's initial position, red plus colour is the desired location of spot, and the magenta cross colour is geometric coherence.

The trajectory path shown in the image of Fig. 8(a) marked with a blue line is the laser beam path followed by the steering laser in an image from the initial position to the desired place at hidden parts of the vocal fold. As expected, the trajectory is straight in the image. Error versus time plot shows an exponential convergence, as shown in Fig. 8(b).

4.2. Stereo-view imaging system and auxiliary mirror simulation result in a realistic case

The second simulation implies a stereoscopic imaging system. Thus, a second camera is added to the first simulation setup, and the control in Eq. (34) is applied. The obtained results in Fig. 9(a) and (b) showed that the laser beam's trajectory path from the initial position to the desired position was straight. Error versus time plot in Fig. 9(c) converged to zero. Similarly, as in Fig. 9(d), mirror velocity had exponential decay.

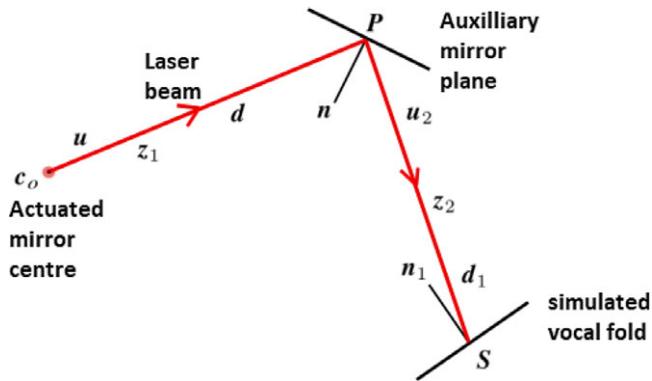


Figure 7. Simulated set-up.

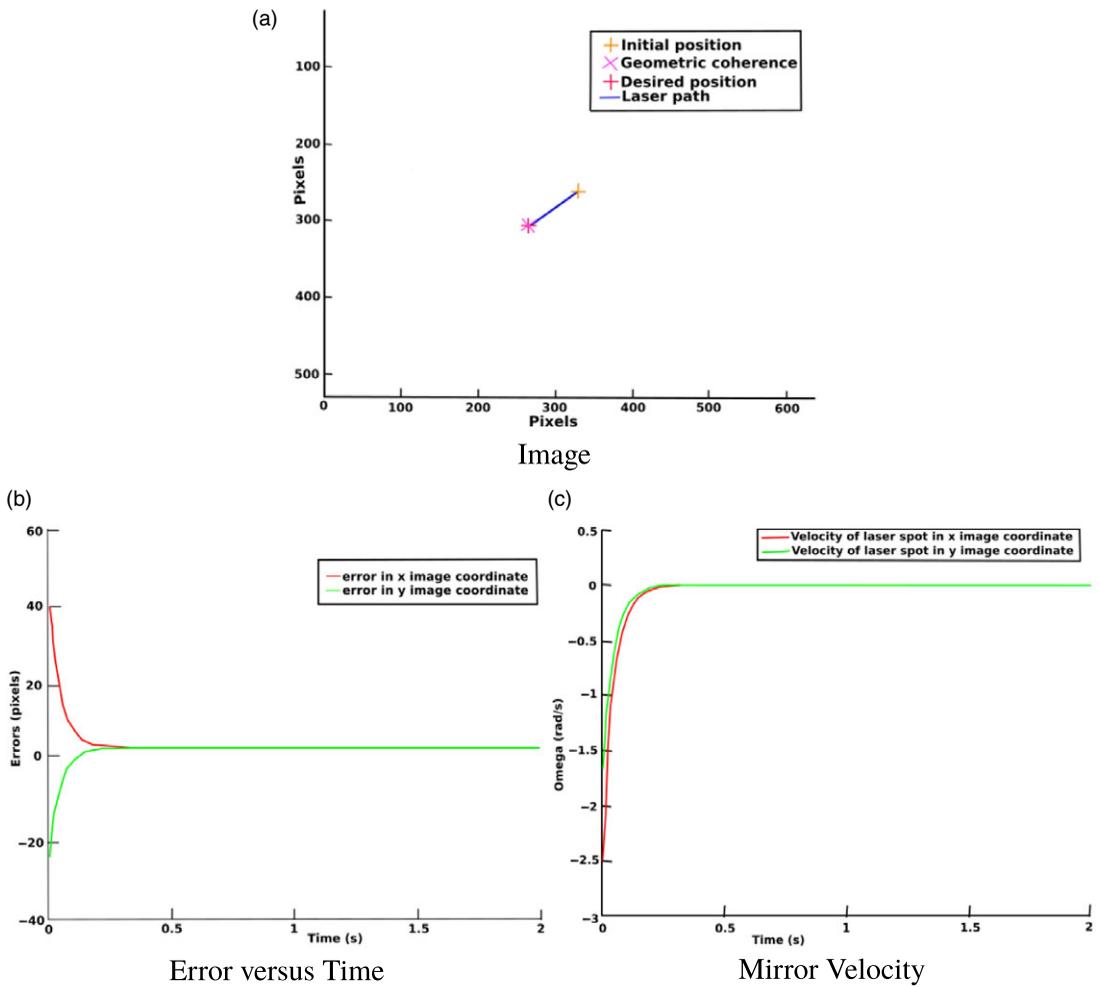


Figure 8. (a) Image. (b) Error versus time. (c) Mirror velocity.

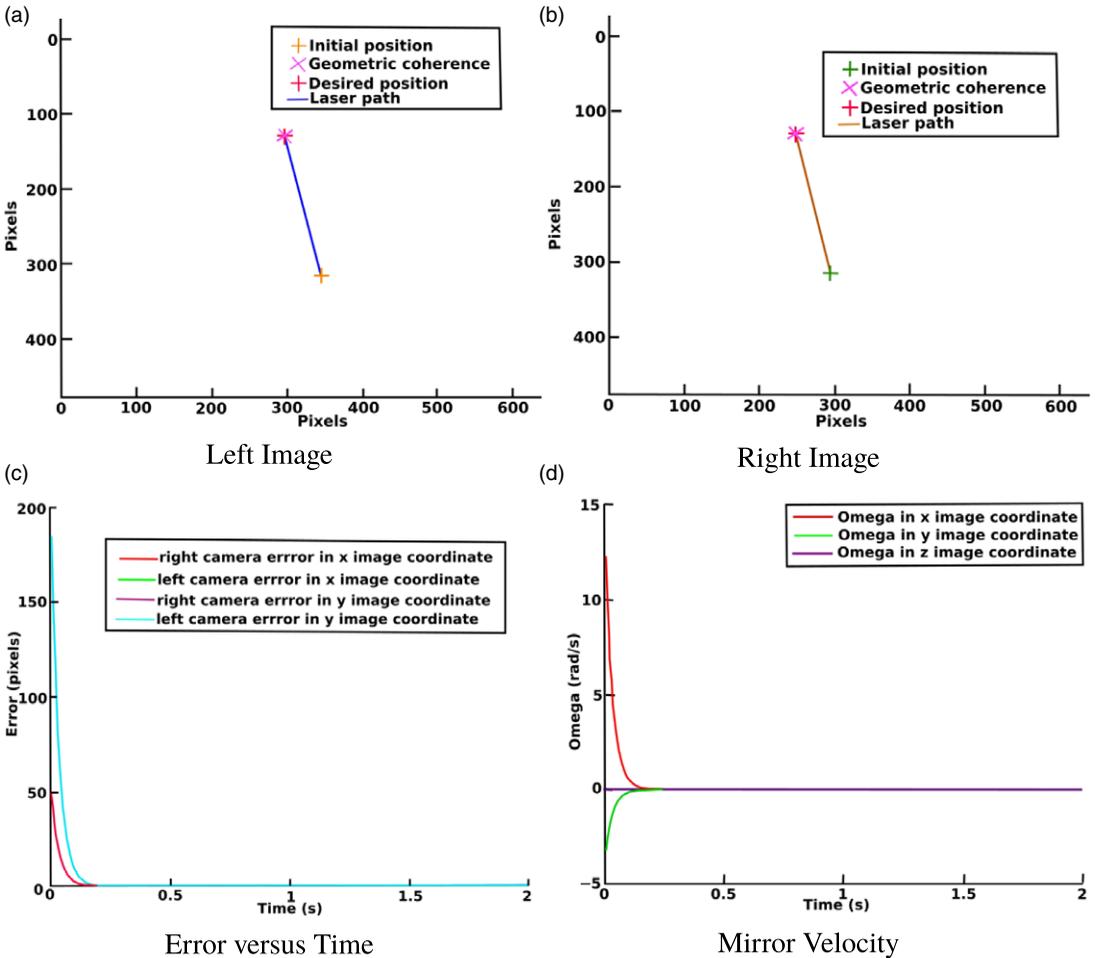


Figure 9. (a) Left image. (b) Right image. (c) Error versus time. (d) Mirror velocity.

5. Experimentation

5.1. Experimental setup

The proposed approaches were validated on the experimental setup shown in Fig. 10. It had two cameras (Guppy pro model: GPF 033B ASG-E0030013 set to work at a frame rate of 25 images per second for a resolution of 640 × 480), a laser source, auxiliary mirror, and an actuated mirror based on a parallel kinematics designed with coplanar axes having triple design, and the platform is driven of three piezo actuators that are located in 120 degrees angles to one another. With the differential drive design, the actuators operate in pairs in a push/pull mode. Two orthogonal rotation axes share a common pivot point. DAC board in a PC generated the analogue input signal; the E-616 PI analogue controller was used to control the actuated mirror’s tip/tilt mirror platform, which caused a change of laser beam direction directed by a laser pointer to the target via an auxiliary mirror. The spot’s position was calculated with respect to the platform centre by a computer (with a program in C++ using the library ViSP <http://visp.inria.fr>). That computed position was sent to a National Instrument card (NI model: USB-6211 with 250 kS/s) externally installed via a USB connection. This communication was enabled using the Labview platform.

The system acquired pixel coordinates through mouse clicks on the image; the user-defined the desired position from which the control algorithm was computed. The corresponding position of the

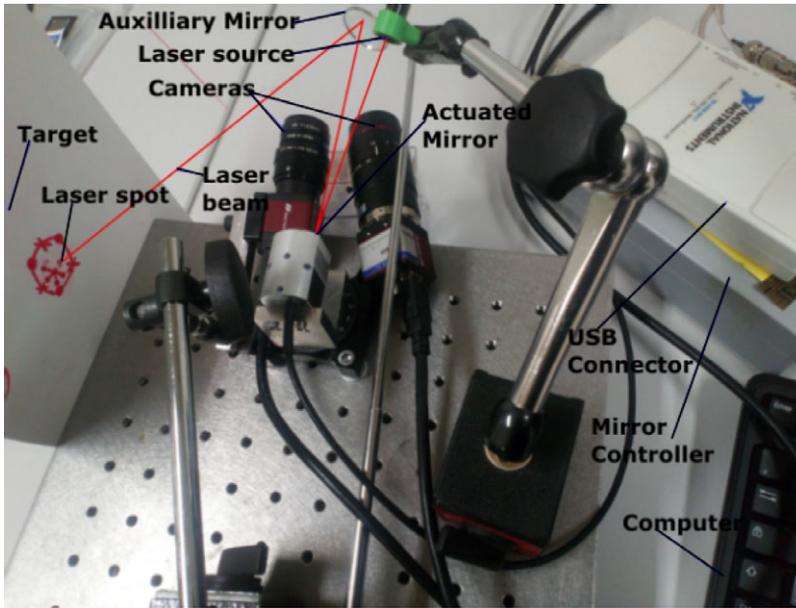


Figure 10. Photography of the experimental setup.

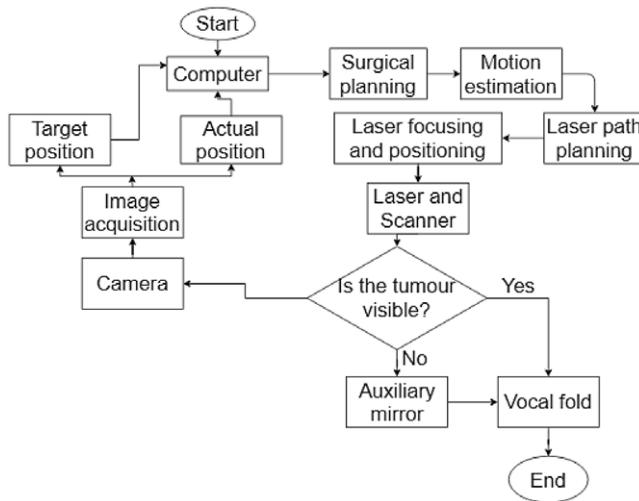


Figure 11. Monocular/ stereoscopic experimental workflow.

micromanipulator thus servoed laser spot from its initial pose to the desired pose. When the laser finally reached its desired position, user-defined through mouse click the next desired position to be attained by the laser.

5.2. Experimental workflow

Figure 11 shows experimental workflow used for both monocular and stereoscopic cases.

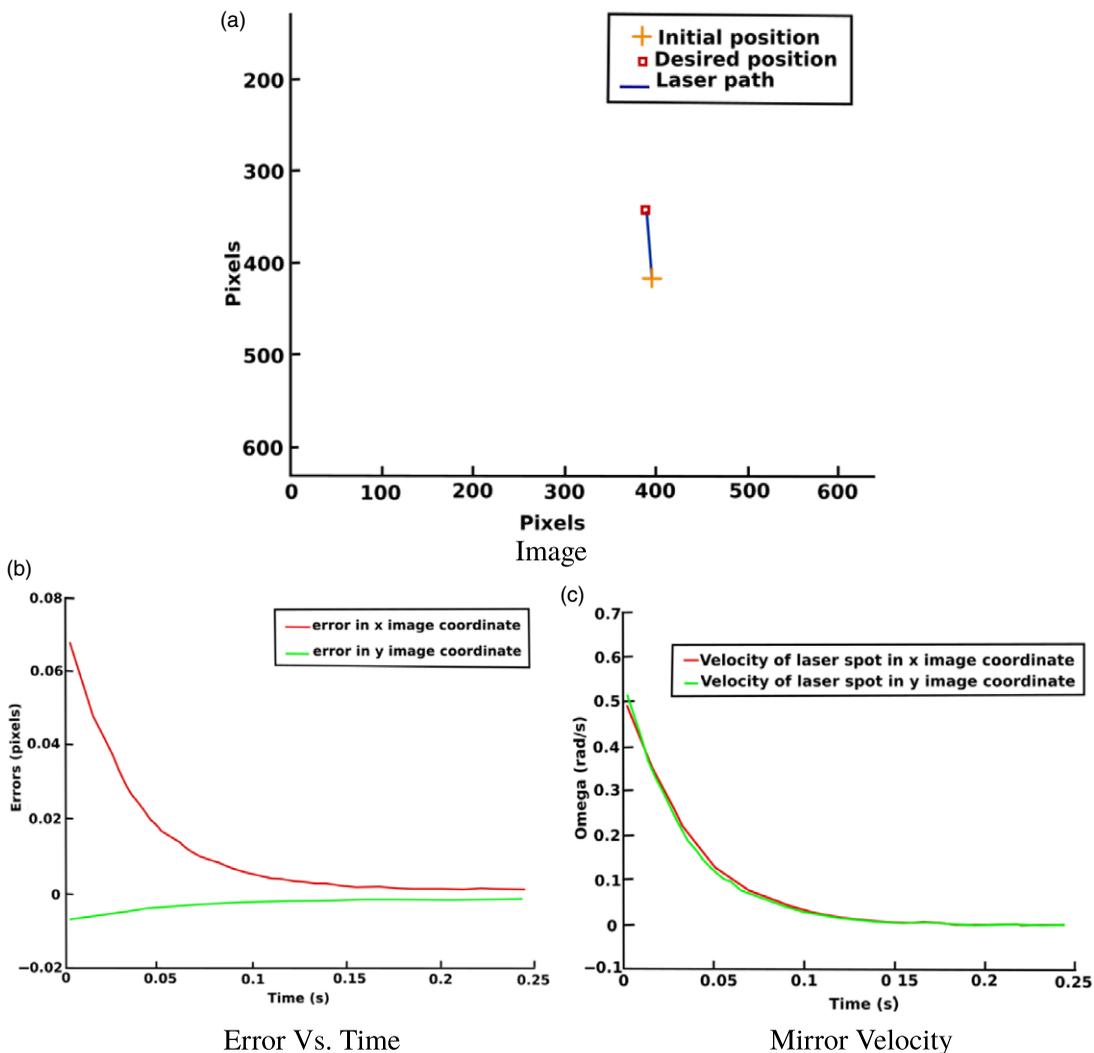


Figure 12. (a) Image. (b) Error versus time. (c) Mirror velocity.

5.3. Single-camera and auxiliary mirror experimental results

Using the setup discussed in Fig. 10, with one camera. The monocular case was validated experimentally, and results obtained in Fig. 12(a) were similar to a simulated case in Section 4.

Figure 12 (b) and (c), error versus time and mirror velocity, respectively, followed exponential decay to reach desired positions, leading to convergence.

5.4. Stereo-view imaging system and auxiliary mirror experimental results

Experimental validation of stereo-view imaging was performed with the setup discussed in Fig. 10. Indeed, Fig. 13(a) and (b) is straight image trajectory path.

Even though both trajectories were straight but for the right image, the path did not reach the desired target; this could be due to laser spot size differences; hence, their centre of gravity moved slightly during control.

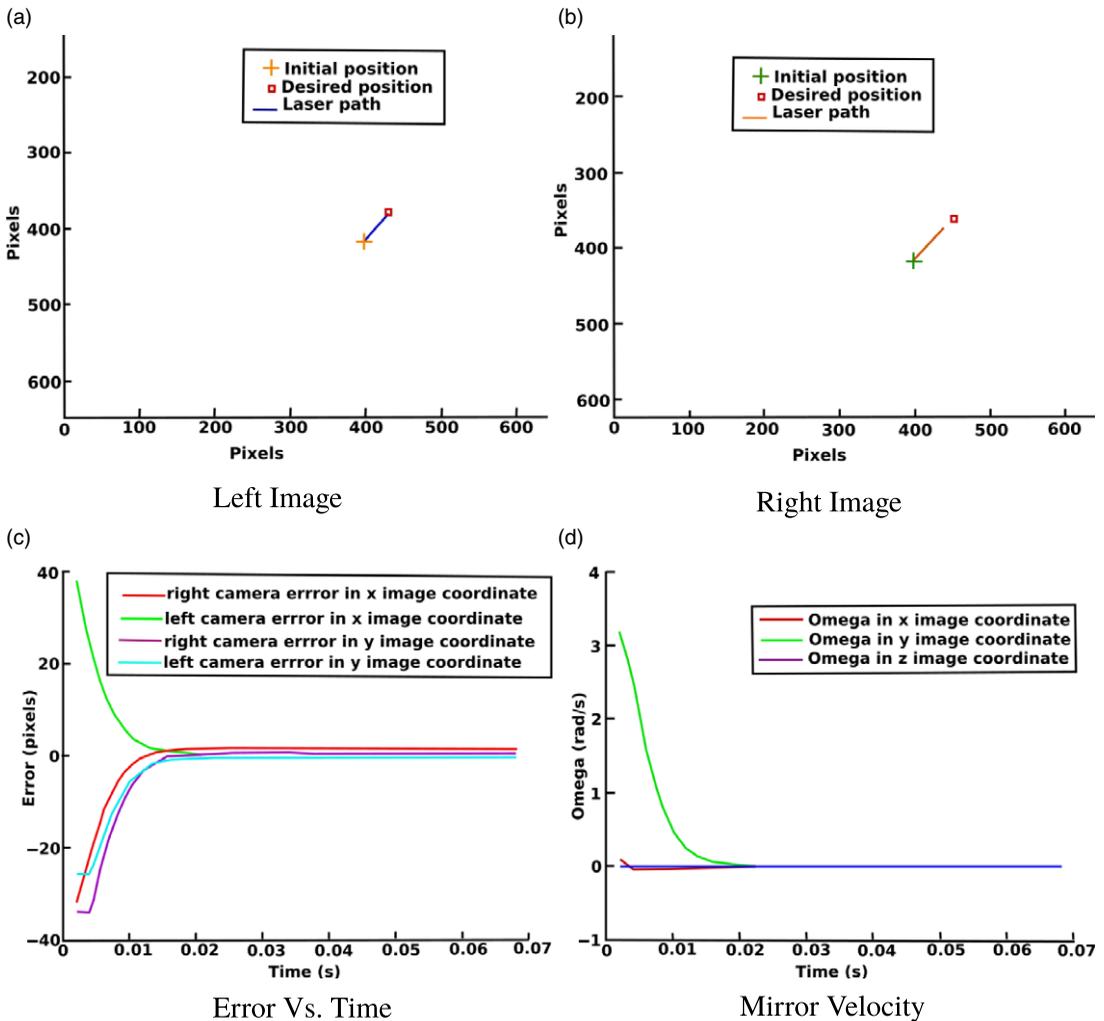


Figure 13. (a) Left image. (b) Right image. (c) Error versus time. (d) Mirror velocity.

Figure 13(c) and (d), error versus time for each image, both *x*, and *y* error components had exponential decay.

Figures 14 and 15 show live video screenshots of laser serving for the conducted experiments.

6. Conclusion

The study shows that vocal fold accessibility improved by seeing through a mirror and serving surgical laser to reach those hidden portions of the vocal fold. Also, the mirror did not affect the controller. The derived control laws could work in both 2D and 3D paths without any prior knowledge of the scene. They were successfully validated in both simulation and experimentally; in all cases, the laser steering control law showed its ability to operate accurately. The experimental results further demonstrated that the proposed control laws were accurate, fully decoupled with exponential decay of the image errors.

The next stages of this study will involve adapting the controller to work under different conditions, for instance, in the influence of perturbations and experimenting on a vocal fold mock-up.

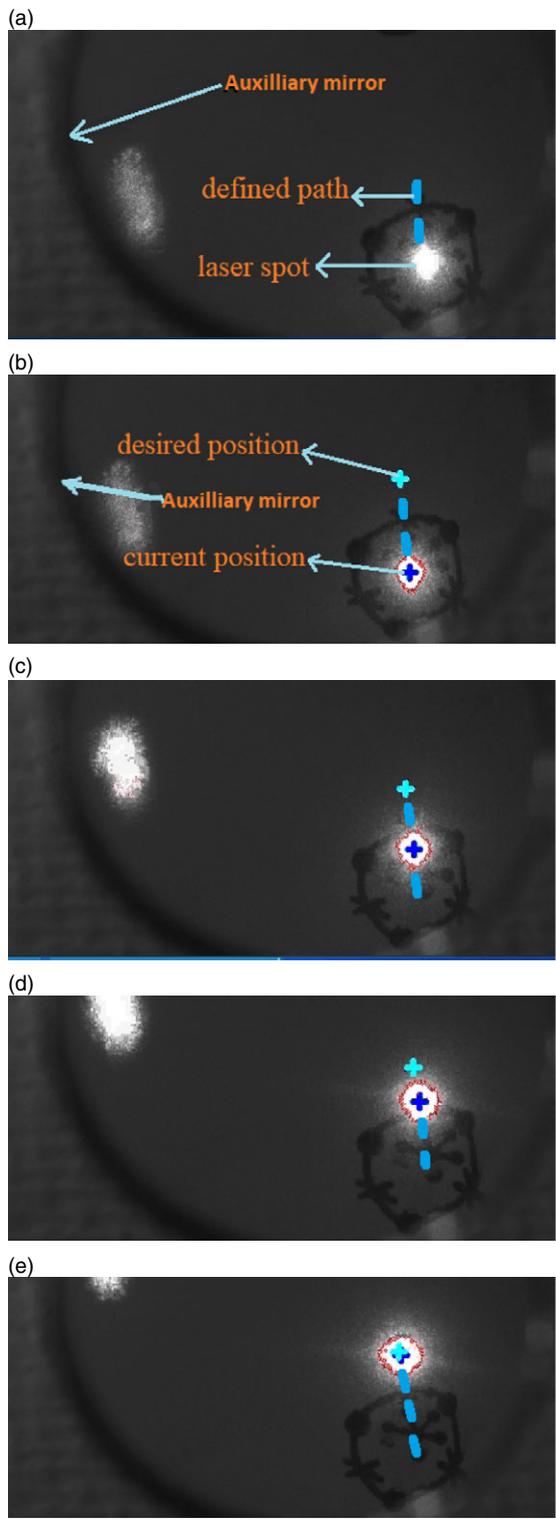
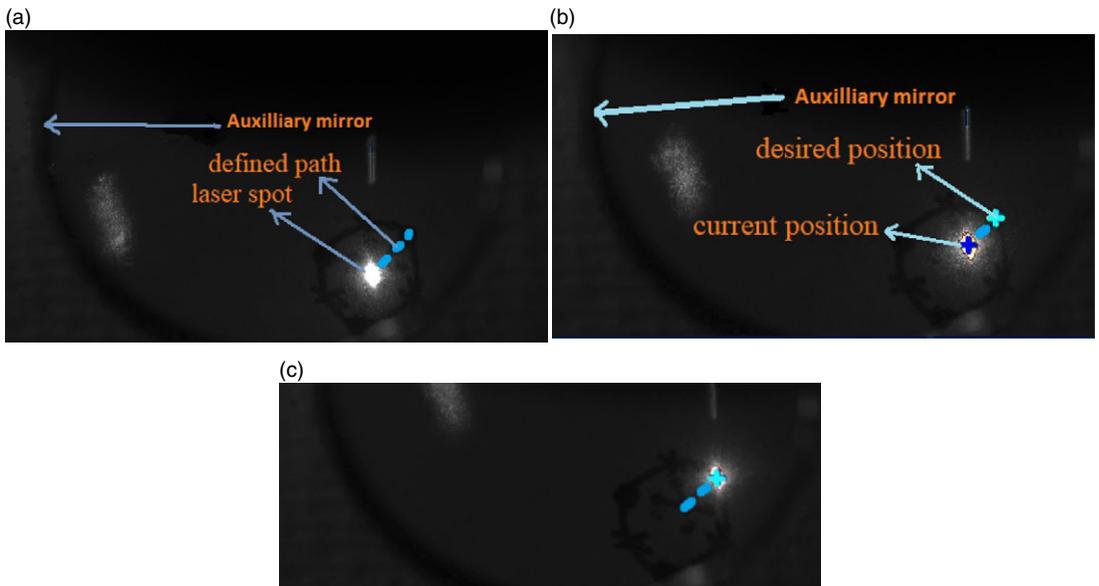


Figure 14. Some experimental images acquired for a single camera system.

Right camera



Left camera

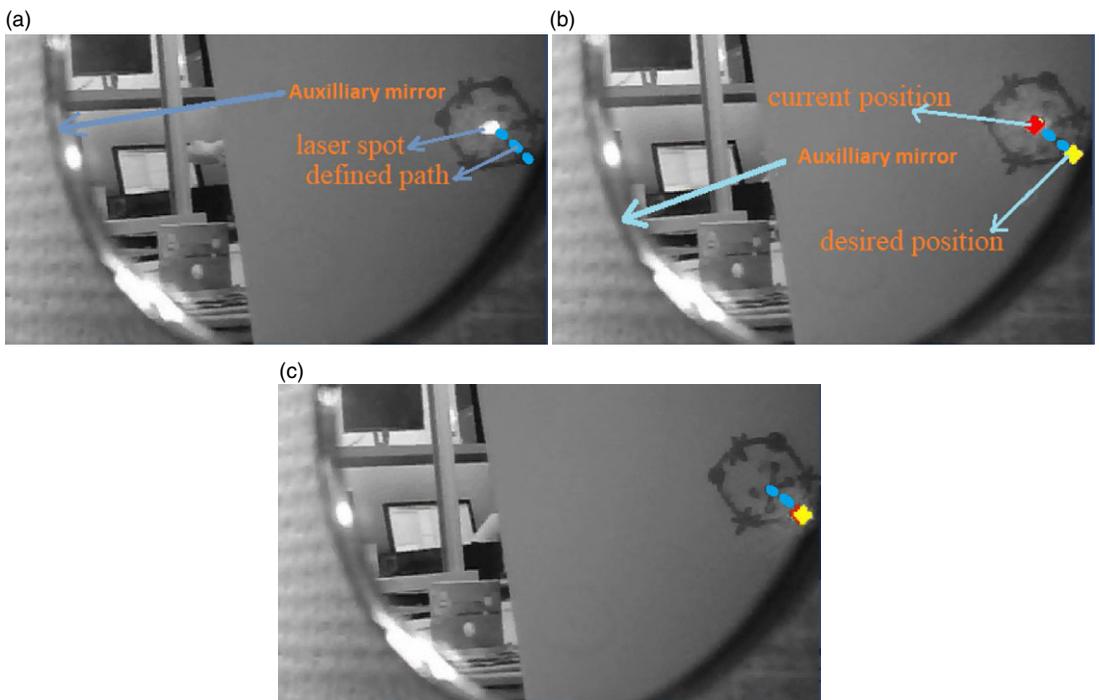


Figure 15. Some experimental images acquired for stereo-view imaging system.

Acknowledgement. This work was supported by the Fondation Charles Defforey – Institut de France through its Grand Prix Scientifique 2018.

Conflicts of Interest. Nicolas Andreff holds a minor share in the capital of Amarob Technologies. All other authors do not have any conflict of interest.

Financial Support. None.

Ethical Considerations. None.

Authors' Contributions. Odira Japheth Ka'pesha is the main author, and all other are coauthors.

References

- [1] S. M. Zeitels, *Atlas of Phonomicrosurgery and Other Endolaryngeal Procedures for Benign and Malignant Disease* (Australia; Great Britain: Singular/Thomson Learning, 2001).
- [2] J. F. II. Giallo, D. Lalush, H. T. Nagle, C. Finley, R. Buckmire, E. Grant, *A medical robotic system for laser phonomicrosurgery* (PhD Dissertation, North Carolina State University, Raleigh, NC, USA, 2008).
- [3] H. E. Eckel, S. Berendes, M. Damm, J. P. Klussmann and K. Wassermann, "Suspension laryngoscopy for endotracheal stenting," *Laryngoscope* **113**(1), 11–15 (2003).
- [4] M. Remacle, "Laser-Assisted Microphonosurgery," **In: Surgery of Larynx and Trachea** (Springer, 2009) pp. 51–56.
- [5] L. Israel, "Digital acublade system. israel@ [online]," (2014). Available: <http://www.surgical.lumenis.com>
- [6] L. S. Mattos and N. Andreff, "The ralp Project: New Technologies and Systems for Robot-Assisted Laser Phonomicrosurgery," *3rd Joint Workshop on New Technologies for Computer/Robot Assisted Surgery, DOI*, vol. **10** (2013).
- [7] R. Renevier, B. Tamadazte, K. Rabenorosoa, L. Tavernier and N. Andreff, "Endoscopic laser surgery: Design, modeling, and control," *IEEE/ASME Trans. Mech.* **22**(1), 99–106 (2016).
- [8] L. S. Mattos, N. Deshpande, G. Barresi, L. Guastini and G. Peretti, "A novel computerized surgeon–machine interface for robot-assisted laser phonomicrosurgery," *Laryngoscope* **124**(8), 1887–1894 (2014).
- [9] N. Andreff and B. Tamadazte, "Laser steering using virtual trifocal visual servoing," *Int. J. Rob. Res.* **35**(6), 672–694 (2016).
- [10] B. Tamadazte and N. Andreff, "Weakly Calibrated Stereoscopic Visual Servoing for Laser Steering: Application to Phonomicrosurgery," *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2014) pp. 743–748.
- [11] S. R. Dukkipati, K.-H. Kuck, P. Neuzil, I. Woollett, J. Kautzner, H. T. McElderry, B. Schmidt, E. P. Gerstenfeld, S. K. Doshi, R. Horton, A. Metzner, A. d'Avila, J. N. Ruskin, A. Natale and V. Y. Reddy, "Pulmonary vein isolation using a visually guided laser balloon catheter: The first 200-patient multicenter clinical experience," *Circ. Arrhythmia Electrophysiol.* **6**(3), 467–472 (2013).
- [12] J.-A. Seon, B. Tamadazte and N. Andreff, "Decoupling path following and velocity profile in vision-guided laser steering," *IEEE Trans. Rob.* **31**(2), 280–289 (2015).
- [13] M. Muddassir, D. Gomez, L. Hu, S. Chen and D. Navarro-Alarcon, "Robotics meets cosmetic dermatology: Development of a novel vision-guided system for skin photo-rejuvenation," *IEEE/ASME Trans. Mech.* (2021). doi: [10.1109/TMECH.2021.3075207](https://doi.org/10.1109/TMECH.2021.3075207)
- [14] B. Espiau, F. Chaumette and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. Rob. Autom.* **8**(3), 313–326 (1992).
- [15] S. Hutchinson, G. D. Hager and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Rob. Autom.* **12**(5), 651–670 (1996).
- [16] A. C. Sanderson and L. E. Weiss, "Image-based Visual Servo Control of Robots," **In: Robotics and Industrial Inspection**, vol. **360** (International Society for Optics and Photonics, 1983) pp. 164–169.
- [17] F. Chaumette and S. Hutchinson, "Visual servoing and visual tracking," **In: Handbook of Robotics** (B. Siciliano and O. Khatib, eds.), Chapter 24, (Springer, 2008) pp. 563–583.
- [18] G. L. Mariottini, S. Scheggi, F. Morbidi and D. Prattichizzo, "Catadioptric Stereo with Planar Mirrors: Multiple-View Geometry and Camera Localization," **In: Visual Servoing via Advanced Numerical Methods** (Springer, 2010) pp. 3–21.
- [19] K. Iida and H. Oku, "Saccade Mirror 3: High-Speed Gaze Controller with Ultra Wide Gaze Control Range Using Triple Rotational Mirrors," *2016 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2016) pp. 624–629.
- [20] K. Okumura, H. Oku and M. Ishikawa, "High-Speed Gaze Controller for Millisecond-Order Pan/Tilt Camera," *2011 IEEE International Conference on Robotics and Automation* (IEEE, 2011) pp. 6186–6191.
- [21] G. L. Mariottini, S. Scheggi, F. Morbidi and D. Prattichizzo, "Planar Catadioptric Stereo: Single and Multi-View Geometry for Calibration and Localization," *2009 IEEE International Conference on Robotics and Automation* (IEEE, 2009) pp. 1510–1515.
- [22] C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 560–576 (2001).

- [23] C. Hue, J.-P. Le Cadre and P. Pérez, “Sequential monte carlo methods for multiple target tracking and data fusion,” *IEEE Trans. Signal Process.* **50**(2), 309–325 (2002).
- [24] S. K. Zhou, R. Chellappa and B. Moghaddam, “Visual tracking and recognition using appearance-adaptive models in particle filters,” *IEEE Trans. Image Process.* **13**(11), 1491–1506 (2004).
- [25] H. T. Nguyen and A. W. Smeulders, “Fast occluded object tracking by a robust appearance filter,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(8), 1099–1104 (2004).
- [26] A. Yilmaz, X. Li and M. Shah, “Contour-based object tracking with occlusion handling in video acquired using mobile cameras,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(11), 1531–1536 (2004).
- [27] Q. Zhao, W. Tian, Q. Zhang and J. Wei, “Robust Object Tracking Method Dealing with Occlusion,” 2016 *3rd International Conference on Soft Computing & Machine Intelligence (ISCFMI)* (IEEE, 2016) pp. 143–147.
- [28] M. I. Shehzad, Y. A. Shah, Z. Mehmood, A. W. Malik and S. Azmat, “K-means based multiple objects tracking with long-term occlusion handling,” *IET Comput. Vision* **11**(1), 68–77 (2017).
- [29] J. Fischer, H. Regenbrecht and G. Baratoff, “Detecting Dynamic Occlusion in Front of Static Backgrounds for AR Scenes,” *Proceedings of the Workshop on Virtual Environments 2003* (2003) pp. 153–161.
- [30] M. M. Wloka and B. G. Anderson, “Resolving Occlusion in Augmented Reality,” *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (1995) pp. 5–12.
- [31] B. Tamadazte, R. Renevier, J.-A. Séon, A. V. Kudryavtsev and N. Andreff, “Laser beam steering along three-dimensional paths,” *IEEE/ASME Trans. Mech.* **23**(30), 1148–1158 (2018).
- [32] E. Marchand and F. Chaumette, “Visual Servoing through Mirror Reflection,” 2017 *IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2017) pp. 3798–3804.