



**HAL**  
open science

# Visual Servoing-based Registration of Multimodal Images

Mohamed Ourak, Brahim Tamadazte, Nicolas Andreff, Eric Marchand

► **To cite this version:**

Mohamed Ourak, Brahim Tamadazte, Nicolas Andreff, Eric Marchand. Visual Servoing-based Registration of Multimodal Images. International Conference on Informatics in Control, Automation and Robotics, Jul 2015, Colmar, France. hal-01159881

**HAL Id: hal-01159881**

**<https://inria.hal.science/hal-01159881>**

Submitted on 4 Jun 2015

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

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

# Visual Servoing-based Registration of Multimodal Images

M. Ourak<sup>1</sup>, B. Tamadazte<sup>1</sup>, N. Andreff<sup>1</sup> and E. Marchand<sup>2</sup>

<sup>1</sup> *FEMTO-ST Institute, AS2M department, Université de Franche-Comté/CNRS/ENSMM, 24 rue Savary, Besançon, France.*

<sup>2</sup> *Université de Rennes 1, IRISA, Rennes, France.  
mouloud.ourak@femto-st.fr, marchand@irisa.fr*

**Keywords:** Visual servoing, Mutual information, Nelder-Mead simplex.

**Abstract:** This paper deals with mutual information-based numerical and physical registration of white light images vs. fluorescence images for microrobotic laser microphonosurgery of the vocal folds. More precisely, it presents two techniques: a numerical registration of multimodal images and a vision feedback control for positioning an endoscope with regards to a preoperative image (fluorescence image). Nelder-Mead Simplex for nonlinear optimization is used to minimize the cost-function. The proposed methods are successfully validated in an experimental set-up using preoperative fluorescence images and real-time white light images of the vocal folds.

## 1 INTRODUCTION

Direct visualization of the larynx and the trachea is often used for the diagnosis but also in surgical intervention (Jackel et al., 2013). The most successful robotic system for the vocal folds surgery is certainly the suspension laryngoscope (Fig. 1(a)). It consists of a straight-rigid laryngoscope, a stereomicroscope, a laser source, and a controlled device based on a foot-pedal activating the laser (Eckel et al., 2003). This system is largely deployed in hospitals but it features many drawbacks: i) extreme extension of the patient's neck; ii) poor ergonomics of the operating setup; iii) considerable skills and expertise required for the clinician; and iv) lack of precision.

Alternative approaches are under investigation: the use of the HARP (Highly Articulated Robotic Probe) highly flexible robot for conventional surgery (Degani et al., 2006) or the use of an endoscopic laser micromanipulator (Tamadazte and Andreff, 2014) (Fig. 1(b)). In both cases, surgery can be preceded by a diagnosis using fluorescence imaging (Sevick-Muraca, 2004). For this, the fluorescence based diagnosis is done a few days before surgery intervention. Therefore, during a surgical intervention the fluorescence diagnosis image must be registered to the real-time white light images grabbed by the endoscopic system in order to define the incision path pour the laser ablation or resection. The Registration can be done either numerically or by physically moving the endoscope to the place where the fluorescence

image was grabbed few days ago.

In this paper, our aim is to control a robot based on direct visual servoing, using image information coming from light white and fluorescence sensors. However, this control needs to be done without a priori model of the robot and the camera. Indeed, approaches have been implemented which are mainly based on the use of the image global information (gradient (Marchand and Collewet, 2010), photometry (Collewet and Marchand, 2011) or mutual information (Dame and Marchand, 2011)). The use of mutual information (MI) in visual servoing problems has proved to be especially effective in the case of multimodal and less contrasted images (Dame and Marchand, 2009). In fact, these control techniques assume that the kinematic model of the robot and the camera intrinsic parameters are at least partially known, but would fail if the system parameters were fully unknown. In addition to the constraint that the initial position cannot be very distant from the desired position to totally ensure convergence. Therefore, it was proposed to use the Simplex method (Nelder and Mead, 1965) instead of gradient (which requires at least a rough calibration of the camera and a computation of the camera/robot transformation) as in (Miura et al., 2005) where the geometrical visual features are used to design the controller.

Furthermore, in the surgical robotics context, it is preferable to free ourselves from any calibration procedure (camera, robot or robot/camera system) for several reasons:

- calibration procedures is often difficult to perform especially by non-specialist operators i.e., clinicians;
- surgeons entering in the operation room are perfectly sterilized to avoid any risk of contamination, and then it is highly recommended to limit the manipulation of the different devices inside the operating room.

For these reasons, we opted for uncalibrated and model-free multimodal registration and visual servoing schemes using the MI as a global visual feature and a Simplex optimization method. This means it is not necessary to compute the interaction matrix (Jacobian image); the kinematic model of the robot may be totally unknown without any constraint in the initial position of the robot with respect to its desired position.

This paper is structured as follows: Section 2 gives the basics of MI. Section 3 presents the Simplex method. Section 4 describes the multimodal registration and the visual servoing achievement. Finally, Section 5 deals with the validation results.

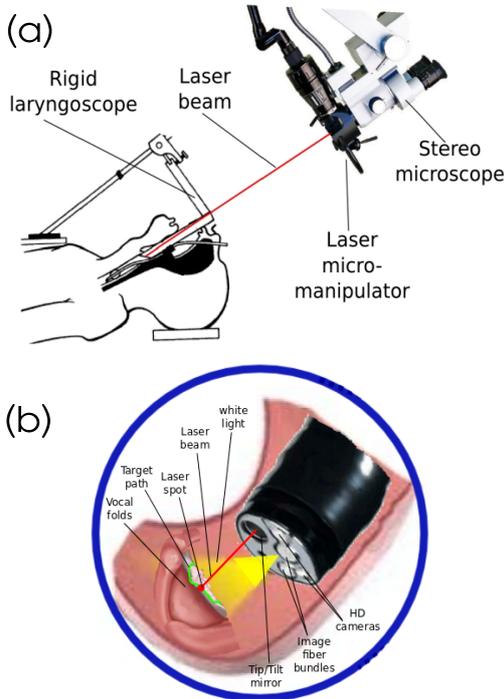


Figure 1: (a) the current laser phonosurgery system and (b) the targeted system.

## 2 MI AND REGISTRATION

In the literature, multimodal images registration has been widely discussed. Zitova et al. (Zitová and Flusser, 2003) classified registration techniques for medical applications into two main categories: area-based and features-based methods. In these cases, the registration process follows mainly four steps: feature detection, feature matching, transformation estimation, and image resampling. As previously stated, our approach is based mutual information rather than geometrical visual features. Therefore, the most critical steps (feature detection and matching) of a conventional registration method are removed. Instead, from the joint and marginal entropy of two images, it is possible to compute their similarities. This means that is the higher the MI, the more the images are aligned (Dame and Marchand, 2009).

### 2.1 Mutual Information in the Image

MI is based on the measure of information, commonly called entropy, in 1D signal. By extension, the entropy expression in an image  $\mathbf{I}$  is given by

$$\mathbf{H}(\mathbf{I}) = - \sum_{i=0}^{N_I} p_{\mathbf{I}}(i) \log_2(p_{\mathbf{I}}(i)) \quad (1)$$

where  $\mathbf{H}(\mathbf{I})$  represents the marginal entropy also called Shannon entropy of an image  $\mathbf{I}$ ,  $i \in [0, N_I]$  (with  $N_I = 255$ ) defines a possible gray value of an image pixel, and  $p_{\mathbf{I}}$  is the probability distribution function also called marginal probability of  $i$ . This can be estimated using the normalized histogram of  $\mathbf{I}$ .

Moreover, the entropy between two images  $\mathbf{I}_1$  and  $\mathbf{I}_2$  is known as joint entropy  $\mathbf{H}(\mathbf{I}_1, \mathbf{I}_2)$ . It is defined as the joint variability of both images

$$\mathbf{H}(\mathbf{I}_1, \mathbf{I}_2) = - \sum_{i=0}^{N_{I_1}} \sum_{j=0}^{N_{I_2}} p_{\mathbf{I}_1, \mathbf{I}_2}(i, j) \log_2(p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)) \quad (2)$$

where  $i$  and  $j$  are the pixels intensities of the two images  $\mathbf{I}_1$  and  $\mathbf{I}_2$  respectively,  $p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)$  is the joint probability for each pixel value. The joint probability is accessible by computing the  $N_{I_1} + 1 \times N_{I_2} + 1 \times N_{bin} + 1$  joint histogram which is built with two axes defining the bin-size representation of the image gray levels and an axis defining the number of occurrences between  $\mathbf{I}_1$  and  $\mathbf{I}_2$ .

From (1) and (2), the MI contained in  $\mathbf{I}_1$  and  $\mathbf{I}_2$  is defined as

$$\mathbf{MI}(\mathbf{I}_1, \mathbf{I}_2) = \mathbf{H}(\mathbf{I}_1) + \mathbf{H}(\mathbf{I}_2) - \mathbf{H}(\mathbf{I}_1, \mathbf{I}_2) \quad (3)$$

and can be expressed using the marginal probability  $p_{\mathbf{I}}$  and joint probability  $p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)$ , by replacing (1)

and (2) in (3) with some mathematical manipulations

$$\mathbf{MI}(\mathbf{I}_1, \mathbf{I}_2) = \sum_{i,j} p_{\mathbf{I}_1, \mathbf{I}_2}(i, j) \log\left(\frac{p_{\mathbf{I}_1 \mathbf{I}_2}(i, j)}{p_{\mathbf{I}_1}(i) p_{\mathbf{I}_2}(j)}\right) \quad (4)$$

which has to be maximized.

In practice, cost-function computed using (4) shows the presence of noise in the MI. This creates more or less important local maxima which complicate the optimization process (Dame and Marchand, 2009). To reduce the joint histogram space as well as MI noise and thereby local maxima (at least for the less significant local maxima), Dawson *et al.* (Dawson and Bowden, 2006) have proposed to use the *in-Parzen* windowing formulation in the MI computation

$$\mathbf{I}_{b1}(k) = \mathbf{I}_1(k) \frac{N_c}{t_{max}} \quad \text{and} \quad \mathbf{I}_{b2}(k) = \mathbf{I}_2(k) \frac{N_c}{t_{max}} \quad (5)$$

where  $t_{max} = r_{max} = 255$  and  $N_c$  the new bin-size of the joint histogram and  $\mathbf{I}_{b1}, \mathbf{I}_{b2}$  is the new gray level value of  $\mathbf{I}_1$  and  $\mathbf{I}_2$ , respectively.

In addition to the resampling of the joint histogram, it is advisable to introduce a filtering method based on *B-splines* interpolation in order to further smooth the MI cost-function. Thus, the abrupt change (mainly due to the fact that we use multimodal images) in the cost-function creating local maxima are *flattened* in order to reduce again these irregularities. In our case, we opted for a third-order interpolation  $\Psi$  which presents a good balance between smoothing quality and time computation. Thereby, both marginal and joint probabilities become

$$p_{\mathbf{I}_{b1} \mathbf{I}_{b2}}(i, j) = \frac{1}{N_k} \sum_k \Psi(i - \mathbf{I}_{b1}(k)) \Psi(j - \mathbf{I}_{b2}(k)) \quad (6)$$

$$p_{\mathbf{I}_{b1}}(i) = \frac{1}{N_k} \sum_k \Psi(i - \mathbf{I}_{b1}(k, x)) \quad (7)$$

$$p_{\mathbf{I}_{b2}}(j) = \frac{1}{N_k} \sum_k \Psi(j - \mathbf{I}_{b2}(k)) \quad (8)$$

with  $N_k$  is the number of pixels in the new images  $\mathbf{I}_{b1}$  and  $\mathbf{I}_{b2}$  and  $\Psi$  is the used B-spline function.

### 3 SIMPLEX-BASED REGISTRATION

This section deals with the method for solving the MI maximization problem. However, before describing the chosen optimization approach among the many existing ones (Kelley, 1999) to solve this problem, it is necessary to know the exact shape of the cost-function in the case of bimodal images (fluorescence *vs.* white light) of the vocal cords.

In practice, rather than maximizing MI, we minimize the cost-function

$$\mathbf{f}(\mathbf{r}) = -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \quad (9)$$

Because the MI depends on a translation of both images, the problem to solve is

$$\hat{\mathbf{r}} = \arg \min_{\mathbf{r} \in SE(3)} \mathbf{f}(\mathbf{r}) \quad (10)$$

where  $\mathbf{r}$  the camera pose with respect to world reference frame.

#### 3.1 Cost-function Shape

Fig. 2 shows the computed cost-function in nominal conditions (i.e., the high definition images shown in Fig. 7). It has a global convex shape but still has many irregularities. Consequently, derivative based methods such as gradient descent could not necessarily guarantee convergence. Thereby, an unconstrained optimization technique was chosen to tackle this problem, i.e., a modified Simplex algorithm.

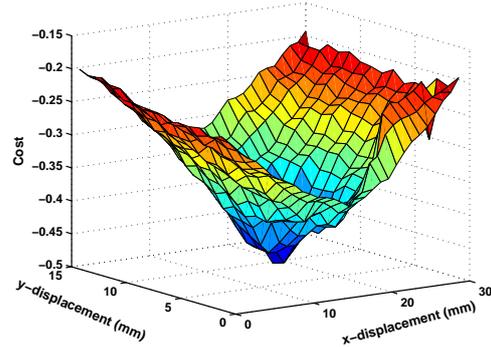


Figure 2: MI cost-function in nominal conditions (representation of  $-\mathbf{MI}$ ).

#### 3.2 Modified Simplex Algorithm

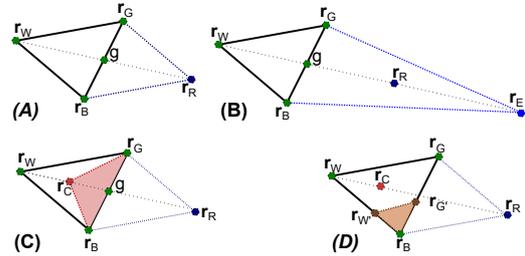


Figure 3: Example of the Simplex steps: (A) reflection, (B) expansion, (C) contraction, and (D) shrinkage.

The Nelder-Mead Simplex algorithm (Nelder and Mead, 1965) roughly work as follows.

A Simplex shape  $S$  defined by vertices  $\mathbf{r}_1 \in SE(3)$  to  $\mathbf{r}_{k+1} \in SE(3)$  with  $k = \dim(6)$  is iteratively updated until convergence using four operators: reflection, contraction, expansion, and shrinkage (see Fig. 3). Thus, the Simplex  $S$  performs the fourth steps as follows

$$\text{reflection: } \mathbf{r}_R = (1 - \alpha)g + \alpha\mathbf{r}_W \quad (11)$$

where  $\mathbf{r}_R$  the reflection vertex,  $\alpha$  the reflection coefficient and  $g$  the centroid between  $\mathbf{r}_G$  and  $\mathbf{r}_B$ .

$$\text{expansion: } \mathbf{r}_E = (1 - \gamma)g + \gamma\mathbf{r}_R \quad (12)$$

where  $\mathbf{r}_E$  the expansion vertex and  $\gamma$  the expansion coefficient, and

$$\text{contraction: } \mathbf{r}_C = (1 - \beta)g + \beta\mathbf{r}_W \quad (13)$$

where  $\mathbf{r}_C$  the contraction vertex, and  $\beta$  the contraction coefficient.

$$\text{shrinkage: } \begin{aligned} \mathbf{r}'_G &= (\mathbf{r}_G + \mathbf{r}_B)/2 \\ \mathbf{r}'_W &= (\mathbf{r}_W + \mathbf{r}_B)/2 \end{aligned} \quad (14)$$

where the vertices are updated as:  $\mathbf{r}_G = \mathbf{r}'_G$  and  $\mathbf{r}_W = \mathbf{r}'_W$ .

Finally, the algorithm ends when  $val(S) \leq \varepsilon$  where  $val(S) = \max(dist(\mathbf{r}_W, \mathbf{r}_B), dist(\mathbf{r}_W, \mathbf{r}_G), dist(\mathbf{r}_G, \mathbf{r}_B))$ ,  $\varepsilon$  is a predefined eligible small distance and  $dist$  is the distance between two vertices.

By convention, the vertices are ordered as

$$f(\mathbf{r}_1) \leq f(\mathbf{r}_2) \leq \dots \leq f(\mathbf{r}_{k+1}) \quad (15)$$

where  $\mathbf{r}_1$  = the best vertex  
 $\mathbf{r}_{k+1}$  = the worst vertex

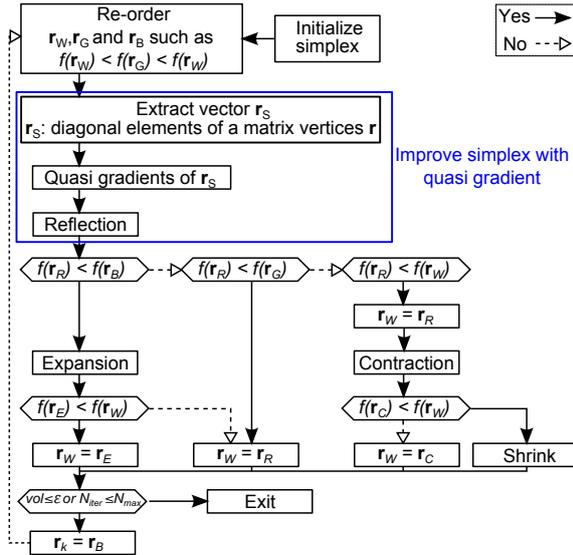


Figure 4: Modified Simplex Algorithm.

The minimization of the cost-function using the Simplex algorithm is shown in Fig. 4. In our case, the Simplex was modified, by introducing the quasi-gradient convergence instead of reflection stage method (Pham and Wilamowski, 2011), in order to improve the convergence direction of  $f$  (without getting trapped in local minima) when the controller approaches the desired position. This combination of an unconstrained and nonlinear method and a quasi-gradient technique allows a higher rate, faster and smooth convergence speed. This returns to combine the advantages of a Simplex and gradient-based optimization methods.

Therefore, (11) is replaced with

$$\mathbf{r}_R = \mathbf{r}_B - \alpha\mathbf{Q} \quad (16)$$

where  $\mathbf{Q}$  the quasi-gradient vector based on the diagonal elements of vertices matrix  $\mathbf{r}$  and  $k$  other points.

## 4 REGISTRATION VS. VISUAL SERVOING

### 4.1 Image Transformation

First, the considered registration is defined as a rigid transformation between two images. Let us assume the transformation  $\hat{\mathbf{r}} \in \mathcal{R}^3 \times SO(1)$  between the white light image  $\mathbf{I}_{b1}$  and the fluorescence image  $\mathbf{I}_{b2}$ . Thereby, this transformation can be estimated by minimizing the value of  $\mathbf{MI}(\mathbf{I}_{b1}, \mathbf{I}_{b2})$ .

$$\hat{\mathbf{r}} = \arg \min -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \mid \mathbf{r} : \mathcal{R}^3 \times SO(1) \quad (17)$$

where  $\mathbf{r}$  a possible rigid transformation. Note that in our case,  $\mathbf{r}$  includes the planar  $xy\theta$  and  $z$  transformations.

The process allowing to carry out this registration is operating as follows: acquisition of both white light image  $\mathbf{I}_{b1}$  and fluorescence image  $\mathbf{I}_{b2}$  then computing  $\mathbf{MI}(\mathbf{I}_{b1}, \mathbf{I}_{b2})$ . The obtained transformation  $\mathbf{r}$  from the first optimization is then applied to synthesize a new image  $\mathbf{I}_{b1}(\mathbf{r})$  from the image  $\mathbf{I}_{b1}$ . These steps are repeated until the predefined stop criterion is reached.

### 4.2 Visual Servoing

Now, it is possible to define a visual servoing control law without any explicit interaction matrix (Jacobian image) or kinematic model robot. Let us assume that we have the cost-function shown in Fig. 2, then our objective is to find the global minimum

$$\hat{\mathbf{r}} = \arg \min_{\mathbf{r} \in SE(3)} -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \quad (18)$$

Note that in our experimental validation, we use a lab-made microrobotic cell having only 3 DOF  $xy\theta$ . This limits, in our case, the transformation to  $\mathbf{r} \in \mathbb{R}^2 \times SO(1)$ .

A first way to move the robot so that the current (smoothed) image  $\mathbf{I}_{b1}$  superimpose onto the desired fluorescence (smoothed) image  $\mathbf{I}_{b2}$  is to use the look-than-move approach: let the Simplex method converge then apply  $\hat{\mathbf{r}}^{-1}$  to the robot and start again (Fig. 6). However, this requires a very fine tuning of the Simplex algorithm. The chosen approach allows interlacing the Simplex loop and the vision-based control loop. At each iteration  $n$ , the Simplex provides  $\mathbf{r}_B^n$ , the best vertex so far, which is associated to the best transformation  ${}^0\mathbf{T}_n$  since the initialization. Thus, applying directly the Simplex would require displacing the robot by

$${}^{n-1}\mathbf{T}_n = {}^0\mathbf{T}_{n-1}^{-1} {}^0\mathbf{T}_n \quad (19)$$

This displacement will not be applied to the complete transformation  ${}^{n-1}\mathbf{T}_n$  found, because that may have the robot to take too large motion. Instead, we extract the screw  $(\delta, \mathbf{t})$  associated to  ${}^{n-1}\mathbf{T}_n$

$${}^{n-1}\mathbf{T}_n = e \begin{pmatrix} [\delta]_{\wedge} & \mathbf{t} \\ 0 & 0 \end{pmatrix} \quad (20)$$

and take only a fraction of it  $\delta = \lambda\delta$  and  $\mathbf{t} = \lambda\mathbf{t}$ , converted to a damped velocity over the sample period  $T_s$  which is  $\omega = \lambda\delta/T_s$  and  $\mathbf{v} = \lambda\mathbf{t}/T_s$ .

Applying this velocity to the robot requires to update the Simplex vertex  $\mathbf{r}_B^n$  according to the current (estimated) transformation (Fig. 5).

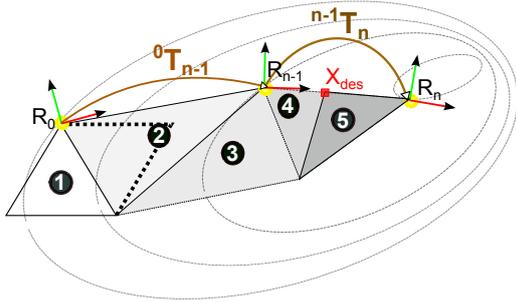


Figure 5: Possible evolution of the Simplex.

$$\mathbf{r}_B^n \Leftrightarrow {}^0\mathbf{T}_n^{update} = {}^0\mathbf{T}_{n-1} e \begin{pmatrix} [\delta]_{\wedge} & \mathbf{t} \\ 0 & 0 \end{pmatrix} \quad (21)$$

This slows down the robot.

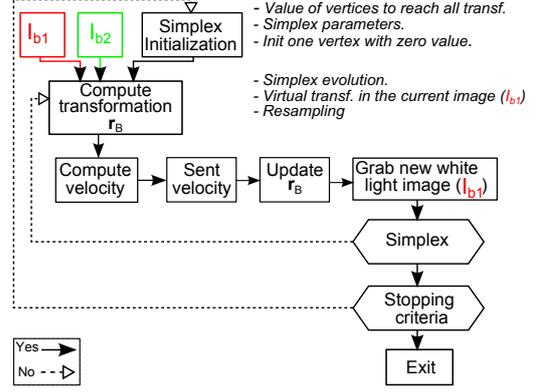


Figure 6: MI-based visual servoing scheme.

## 5 REAL-WORLD VALIDATION

### 5.1 Numerical Registration Results

The proposed numerical registration method is validated using two vocal folds images: real fluorescence and white light. These images taken from (Arens et al., 2004) were acquired in two different points of view with known pose as shown in Fig. 7. It can be highlighted that  $\hat{\mathbf{r}}$  between  $\mathbf{I}_{b1}$  and  $\mathbf{I}_{b2}$  includes four parameters ( $x, y, \theta$  and  $zoom$ ). To be more realis-

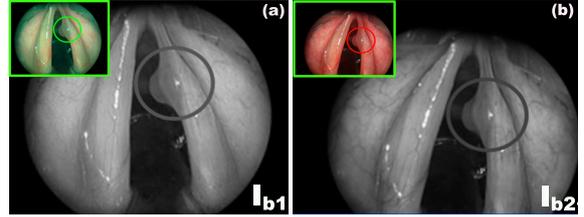


Figure 7: (a) fluorescence image  $\mathbf{I}_{b2}$  and (b) white light image  $\mathbf{I}_{b1}$ .

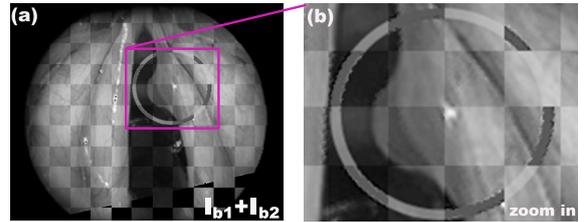


Figure 8: Numerical registration results: (a) shows  $\mathbf{I}_{b1}$  integrated in  $\mathbf{I}_{b2}$ , and (b) a zoom in of region of interest.

tic in our validation tests, we added a circular trajectory (i.e., virtual incision mark done by a surgeon), to be tracked by the surgical laser spot, in the fluorescence image delimiting the tumor (Fig. 7). This path is then by analyzing Fig. 8(a), it can be underlined the continuity of the combination  $\mathbf{I}_{b1} + \mathbf{I}_{b2}$  which

explains the high accuracy of the registration method, this is clearly visible on the zoom in the incision mark (Fig. 8(b)). For this example, the numerical values are summarized in Table 1.

Table 1: Numerical values of  $\hat{\mathbf{r}}, \hat{z}$  ( $1pix = 0.088mm$ ).

DOF	real pose	obtained pose	errors
$x$ (mm)	-8.000	-7.767	0.233
$y$ (mm)	-12.000	-12,059	0.059
$\theta$ (deg)	12.000	12.500	0.500
$z$	1.09	1.089	0.010

## 5.2 Visual Servoing Results

It is not yet possible to test our developments directly on true vocal folds using the developed flexible laryngoscope by current lack of a technical solution for fiber-based surgical laser and fiber-based imaging system (this problem is being addressed through the EU project  $\mu$ RALP: <http://www.microralp.eu>) (see Fig.14). Therefore, a 3 DOF ( $xy\theta$ ) microrobotic cell is used for the validation (Fig. 9).

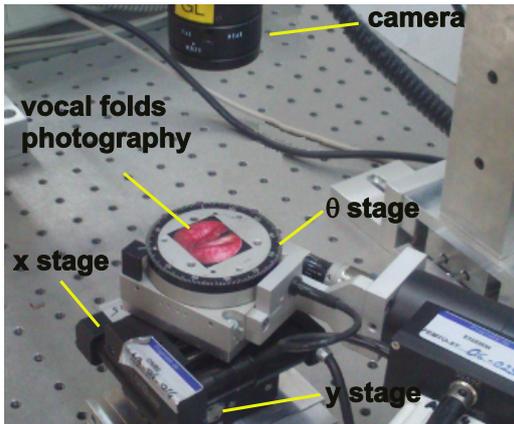


Figure 9: Photography of the microrobotic cell.

Firstly, the MI-based visual servoing is validated on monomodal images in aim to verify the validity of our controller. Fig. 10 represents an example of white light images registration in visual servoing mode. More precisely, Fig. 10(a) and (b) represent the initial and desired images, respectively. In the same way, Fig. 10(c) and (d) show the initial and final error  $\mathbf{I}_{b1} - \mathbf{I}_{b2}$ . It can be noticed that the final position of the positioning platform matches perfectly with the desired position indicating the good accuracy of our method.

Fig. 11 shows the evolution of the velocities  $v_x$ ,  $v_y$  and  $v_\theta$  in the different DOF versus number of iterations  $i$ . It can be underlined that the developed

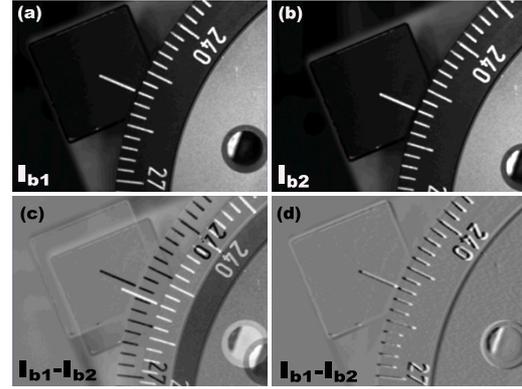


Figure 10: MI-based visual servoing (white light images).

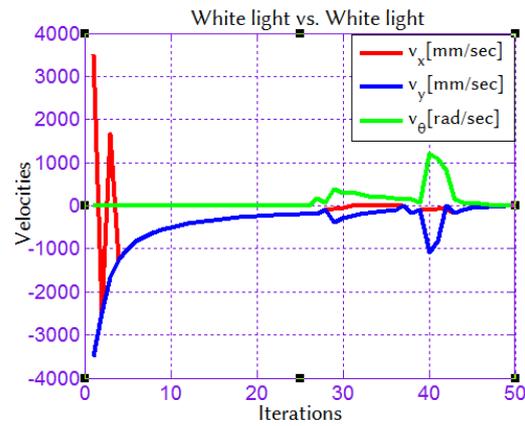


Figure 11: Image velocities  $v_x$ ,  $v_y$  and  $v_\theta$  vs. iterations  $i$ .

controller converges with accuracy in fifty iterations (each iteration takes about 0.5 second). Also, the speed varies in the iteration 40 because the Simplex after initialization found a new best minimum.



Figure 12: MI-based visual servoing (white light vs. fluorescence images).

Secondly, vocal folds multimodal images are also used to test the proposed controller. In this scenario, the desired image is in fluorescence mode (pre-

corded image) and the current images are in white light mode as it would be in the surgical context. Fig. 12(a) and (b) show the initial image  $\mathbf{I}_{b1}$  and the desired image  $\mathbf{I}_{b2}$ , respectively. Fig. 12(c) and (d) illustrate the error  $\mathbf{I}_{b1} - \mathbf{I}_{b2}$  during the visual servoing process. As shown in this figure, the controller converges also to the desired position with a good accuracy. Note that the image  $\mathbf{I}_{b1} - \mathbf{I}_{b2}$  is not completely gray (if two pixels are exactly the same, it is assigned the gray value of 128 for a better visualization of  $\mathbf{I}_{b1} - \mathbf{I}_{b2}$ ), this is due to the fact that both images are acquired from two different modalities, then the difference will never be zero (respectively 128 in our case).

In the same manner, Fig. 13 shows the evolution of the velocities  $v_x$ ,  $v_y$  and  $v_\theta$  with respect number of iterations  $i$ . It can be also underlined that the controller converges with accuracy to the desired position despite the large difference between  $\mathbf{I}_{b1}$  and  $\mathbf{I}_{b2}$ .

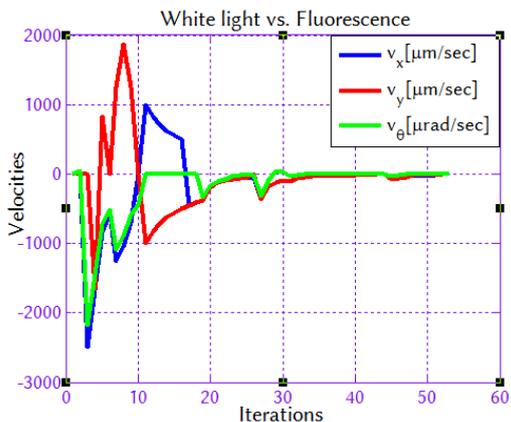


Figure 13: Image velocities  $v_x$ ,  $v_y$  and  $v_\theta$  vs. iterations  $i$ .

Additional validation tests were performed to assess the repeatability and behavior (convergence and robustness) of the controller. Therefore, for each test, the experimental conditions (lighting conditions, initial position and image quality) were deliberately altered. Table 2 gives the results of a sample of these experiments.

## 6 CONCLUSION

In this paper, a novel metric visual servoing-based on MI has been presented. Unlike the traditional methods, the developed approach was based only on the use of a modified Simplex optimization without any knowledge of neither robot, camera models nor camera/robot transformation. It has been shown that the designed controller works even in the presence of

Table 2: Repeatability test for visual servoing ( $x$ ,  $y$ , and  $\theta$  in  $^\circ$ ) and  $t$  in seconds.

N <sup>o</sup>	DOF	des. pos.	ini. pos.	error	$t$
1	$x$	5.37	2.47	-0.33	25.2
	$y$	2.94	0.66	0.37	
	$\theta$	-2.61	-8.43	2.51	
2	$x$	4.02	-0.66	0.37	36.5
	$y$	-5.57	-5.05	1.45	
	$\theta$	2.47	-5.05	2.41	
3	$x$	6.05	3.14	0.16	49.2
	$y$	1.47	0.21	0.88	
	$\theta$	-14.59	-24.19	0.64	
4	$x$	4.09	2.1	0.17	36.3
	$y$	2.12	0.44	0.4	
	$\theta$	14.56	6.63	1.15	
5	$x$	3	0.31	0.55	57.3
	$y$	2.5	0.19	0.53	
	$\theta$	-4.81	14.53	0.83	

many local minima in the MI cost-function. Beside this, the controller has shown good behavior in terms of accuracy, repeatability and convergence.

Future work will be devoted to optimize the computation time to reach the video rate and implement this approach in the developed laser vocal folds surgery system shown in Fig 14. Also, we will discuss the validation of the controller in SE(3) using a 6 DOF robot.

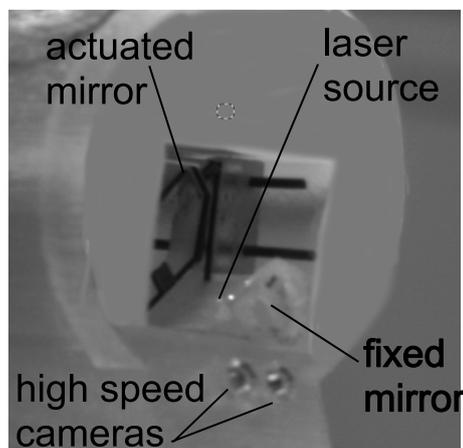


Figure 14: Photography of the developed endoscopy tip.

## ACKNOWLEDGEMENTS

This work was supported by  $\mu$ RALP, the EC FP7 ICT Collaborative Project no. 288663 (<http://www.microralp.eu>), and by ACTION, the

French ANR Labex no. ANR-11-LABX-01-01  
(<http://www.labex-action.fr>).

## REFERENCES

- Arens, C., Dreyer, T., and Glanz, H. Malzahn, K. (2004). Indirect autofluorescence laryngoscopy in the diagnosis of laryngeal cancer and its precursor lesions. *European Archives of Oto-Rhino-Laryngology and Head & Neck*, 261(2):71–76.
- Collewet, C. and Marchand, E. (2011). Photometric visual servoing. *IEEE Trans. on Robotics*, 27(4):828–834.
- Dame, A. and Marchand, E. (2009). Entropy-based visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, pages 707–713, Kobe, Japan.
- Dame, A. and Marchand, E. (2011). Mutual information-based visual servoing. *IEEE Trans. on Robotics*, 27(5):958–969.
- Degani, A., Choset, H., Wolf, A., and Zenati, M. A. (2006). Highly articulated robotic probe for minimally invasive surgery. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 4167–4172. IEEE.
- Dowson, N. and Bowden, R. (2006). A unifying framework for mutual information methods for use in non-linear optimisation. In *Computer Vision ECCV*, volume 3951 of *Lecture Notes in Computer Science*, pages 365–378. Springer Berlin Heidelberg.
- Eckel, H., Berendes, S., Damm, M., and Klusmann, J. (2003). Suspension laryngoscopy for endotracheal stenting. *Laryngoscope*, 113:11–15.
- Jackel, M., Martin, A., and Steine, W. (2013). Twenty-five years experience with laser surgery for head and neck tumors. volume 264 of *European Archives of Oto-Rhino-Laryngology*, pages 577–585.
- Kelley, C. (1999). *Iterative Methods for Optimization*. Frontiers in Applied Mathematics 18.
- Marchand, E. and Collewet, C. (2010). Using image gradient as a visual feature for visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'10*, pages 5687–5692, Taipei, Taiwan.
- Miura, K., Hashimoto, K., Gangloff, J., and de Mathelin, M. (2005). Visual servoing without jacobian using modified simplex optimization. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 3504–3509. IEEE.
- Nelder, A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.
- Pham, N. and Wilamowski, B. (2011). Improved nelder mead’s simplex method and applications. *Journal of computing*, 3(3):55–63.
- Sevick-Muraca, E. (2004). Fluorescence-enhanced optical imaging and tomography for cancer diagnostics. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, volume 2, pages 1482–1485.
- Tamadazte, B. and Andreff, N. (2014). Weakly calibrated stereoscopic visual servoing for laser steering: Application to phonomicrosurgery. In *IROS*, pages 743–748.
- Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.