Multimodal Image Registration and Visual Servoing

M. Ourak[1], B. Tamadazte[1], N. Andreff[1] and E. Marchand[2]

[1] FEMTO-ST Institute, AS2M department, Univ. Bourgogne Franche-Comté/CNRS/ENSMM, 24 rue Savary, F-25000 Besançon, France.
[2] Université de Rennes 1, IRISA, Rennes, France.
mouloud.ourak@femto-st.fr

Abstract. This paper deals with multimodal imaging in the surgical robotics context. On the first hand, it addresses numerical registration of a preoperative image obtained by fluorescence with an intraoperative image grabbed by a conventional white-light endoscope. This registration involves displacement and rotation in the image plane as well as a scale factor. On the second hand, a method is developped to visually servo the endoscope to the preoperative imaging location. Both methods are original and dually based on the use of mutual information between a pair of fluorescence and white-light images and of a modified Nelder-Mead simplex algorithm. Numerical registration is validated on real images whereas visual servoing is validated experimentally in two set-ups: a planar microrobotic platform and a 6DOF parallel robot.

1 Introduction

This work is grounded into robot assisted laser phonosurgery (RALP). The current gold standard procedure for the vocal folds surgery is certainly suspension microlaryngoscopy (Fig. 1(a)) which requires direct visualization of the larynx and the trachea as proposed in [9]. This system is widely deployed in hospitals but it features many drawbacks related to patient and staff safety and comfort. Therefore, alternative endoscopic approaches are under investigation: the extended use of the HARP (Highly Articulated Robotic Probe) highly flexible robot, designed for conventional surgery [6] or the use of an endoscopic laser micro-manipulator [16] (Fig. 1(b)). In all aforementioned cases, cancer diagnosis can be performed thanks to fluorescence imaging [15], (a few) days before the surgical intervention. The latter is usually performed under white-light conditions because fluorescence may require longer exposure time than real time can allow. 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 of the laser ablation or resection. Registration can be done either numerically or by physically servoing the endoscope to the place where the preoperative fluorescence image was grabbed.

In this paper, our aim is to control a robot based on direct visual servoing, *i.e.* using image information coming from white light and fluorescence sensors.

Several visual servoing approaches based on the use of features (line, Region of interest (ROI)) [2] or the image global information (gradient [11], photometry [3] or mutual information [5]) can be used. Nevertheless, 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 [4]. 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 practice, the initial position cannot be very distant from the desired position to ensure convergence. To enlarge the stability domain, [12] proposed to use the Simplex method [13] instead of the usual gradient-like methods (which require at least a rough calibration of the camera and a computation of the camera/robot transformation). However, the work in [12] relies on the extraction from the image of geometrical visual features.

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:

1. Calibration procedures are often difficult to perform, especially by non-specialist operators i.e., clinicians.
2. Surgeons entering in the operating 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 mutual information as a global visual feature and a Simplex as optimization approach. Thereby, 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. A preliminary version of this work was presented in [19] in the case of planar positioning and is extended in this paper to positioning in the 3D space.

This paper is structured as follows: Section 2 explains the medical application of the proposed approach. Section 3 gives the basic background on mutual information. Section 4 presents a modified Simplex method. Section 5 describes multimodal registration and multimodal visual servoing. Finally, Sec. 6 deals with the validation results.

2 Medical Application

The vocal folds are situated at the center and across the larynx and form a V-shaped structure. They are used to create the phonation by modulating the air flow being expelled from the lungs through quasi-periodic vibrations. They can be affected by benign lesions, such as cysts or nodules (for instance, when they are highly stressed, *e.g.* when singing) or, in the worst case, cancer tumors (especially for smokers). These lesions change the configuration of the folds and thereby the patient's voice. Nowadays, medical tools can be used to suppress

this trouble and recover the original voice in particular for cyst and nodules. Appeared in 1960, *phonosurgery* – the surgery of the vocal folds – can be divided into laryngoplastic, laryngeal injection, renovation of the larynx and phonomicrosurgery. Specifically, laser phonomicrosurgery consists of a straight rigid laryngoscope, a stereoscopic microscope, a laser source, and a controlled 2DOF device to orient the laser beam [8], as shown in Fig. 1(a). Nevertheless, the current system requires extreme skill from the clinician. Specifically, high dexterity is required because both the laser source is located out of the patient, 400 mm away from the vocal folds. This distance increases the risk of inaccuracy when the laser cutting process is running. Moreover, the uncomfortable position of the patient's neck in a straight position all along the operation can be traumatic. The drawbacks of the conventional procedure are taken into account in

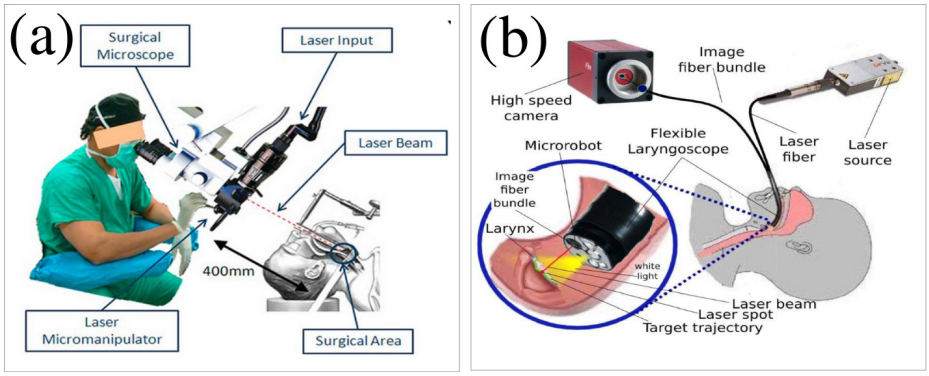

Fig. 1. Global view of the microphonosurgery system: (a) the current laser microphonosurgery system and (b) the targeted final system.

the new set-up developed within the European project μRALP, which consists on embedding all the elements (i.e., cameras, laser and guided mirror) inside an endoscope Fig. 1(b). More precisely, the endoscope is composed of white light, high speed camera imaging the laser evolution with 3D feedback to the clinician. Additionnally, a low framerate, high sensitivity fluorescence imaging system is to be used preoperatively to detect cancerous lesions.

The global approach is based on the use of 2 degrees of freedom (DOF) to guide the laser along the trajectory drawn by the surgeon on a preoperative fluorescence image. However, since the preoperative image is not necessarily taken by the same instrument on the same location, this approach requires the preoperative fluorescence image (where the surgeon decides the trajectory) and the white light image (where the control of the robot is developed) to be registered. This can be done in two ways: registration or servoing. Registration deals with the estimation of the transformation between both images, which can then be used to morph the fluorescence image onto the real-time endoscopic image flow

(for instance, as an augmented reality). Visual servoing deals with bringing the endoscope back to the place where the fluorescence image was grabbed and stabilizing it in that configuration, which amounts to a physical registration and should turn useful in many other applications, such as surgery in the stomach to compensate for physiological motions.

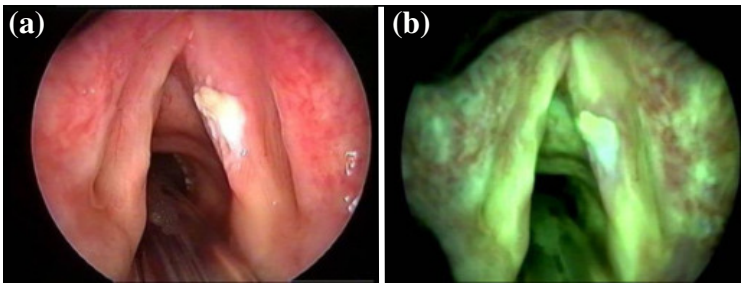

Fig. 2. Vocal folds endoscopic images: (a) white light endoscopic image (b) fluorescence endoscopic image [17].

3 Mutual Information and Registration

In the literature, multimodal image registration has been widely discussed. Zitova et *al.* [18] 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 on 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 the higher the mutual information is, the better the images are aligned [4].

3.1 Mutual Information in the Image

Mutual information 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

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

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

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

where i and j are the pixel intensities of the two images $\mathbf{I}_1$ and $\mathbf{I}_2$ respectively; and $p_{\mathbf{I}_1\mathbf{I}_2}(i,j)$ is the joint probability for each pixel value. The joint probability is accessible by computing the $(N_{\mathbf{I}_1}+1) \times (N_{\mathbf{I}_2}+1) \times (N_{bin}+1)$ 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 mutual information contained in $\mathbf{I}_1$ and $\mathbf{I}_2$ is defined as

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

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

$$\mathbf{MI}(\mathbf{I}_1, \mathbf{I}_2) = \sum_{i,j} p_{\mathbf{I}_1,\mathbf{I}_2}(i,j) log\Big(\frac{p_{r\mathbf{I}_1\mathbf{I}_2}(i,j)}{p_{\mathbf{I}_1}(i) p_{\mathbf{I}_2}(j)}\Big) \tag{4}$$

This cost-function has to be maximized if $\mathbf{I}_1$ and $\mathbf{I}_2$ are requested to "look like each other".

In practice, the cost-function computed using (4) is not very smooth. This creates local maxima, hence complicating the convergence optimization process [4]. To reduce the joint histogram space as well as the irregularities in the mutual information, and thereby local maxima (at least for the less significant ones), Dawson et *al.* [7] proposed to use the *in-Parzen* windowing formulation when computing the mutual information:

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

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

In addition to re-sampling of the joint histogram, it is advisable to introduce a filtering method based on *B-splines* interpolation in order to further smooth the mutual information cost-function. As far as multimodal images are concern, the abrupt change in the cost-function creating local maxima are *flattened* in order to reduce again these irregularities. In practice, we opted for a third-order interpolation ψ, which presents a good balance between smoothing quality and

time computation. Thereby, both marginal and joint probabilities become

$$p_{\mathbf{I}_{b1}\mathbf{I}_{b2}}(i,j) = \frac{1}{N_k}\sum_k \psi\left(i - \mathbf{I}_{b1}(k)\right)\psi\big(j - \mathbf{I}_{b2}(k)\big) \tag{6}$$

$$p_{\mathbf{I}_{b1}}(i) = \frac{1}{N_k}\sum_k \psi\left(i - \mathbf{I}_{b1}(k,x)\right) \tag{7}$$

$$p_{\mathbf{I}_{b2}}(j) = \frac{1}{N_k}\sum_k \psi\left(j - \mathbf{I}_{b2}(k)\right) \tag{8}$$

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

4 Simplex-based Registration

This section deals with the method for solving the mutual information maximization problem. However, before describing the chosen optimization approach among the many existing ones [10] 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 mutual information, we minimize the cost-function

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

In the general case, because the mutual information depends on a Euclidean displacement (*i.e.* in $SE(3)$) between both image viewpoints, the problem to solve is

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

where $\mathbf{r}$ is the camera pose with respect to the world reference frame, attached to the fluorescence image.

4.1 Cost-function Shape

Figure 3 shows the computed cost-function in nominal conditions (i.e., the high definition images shown in Fig. 8). 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 overcome this problem, i.e., a modified Simplex algorithm.

4.2 Modified Simplex Algorithm

The Nelder-Mead Simplex algorithm [13] roughly works as follows. A Simplex shape S defined by vertices $\mathbf{r}_1$ to $\mathbf{r}_{k+1}$ with $k = dim(6)$ is iteratively updated

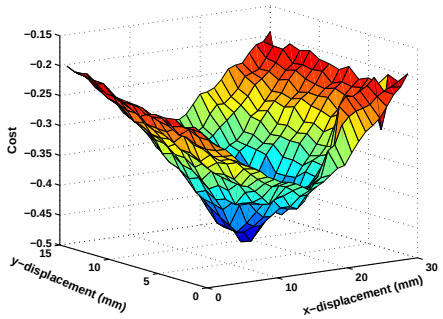

Fig. 3. MI cost-function in nominal conditions (representation of -MI).

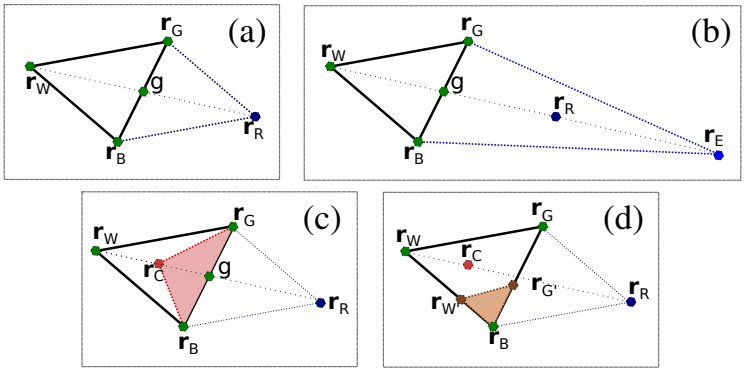

Fig. 4. Example of the Simplex steps: (A) reflection, (B) expansion, (C) contraction, and (D) shrinkage.

until convergence using four operators: reflection, contraction, expansion, and shrinkage (see Fig. 4), defined on a linear space.

In order to apply this algorithm in the non linear Euclidean space, we represent any rigid displacement $\mathbf{r} \in SE(3)$ as

$$\mathbf{r} = \begin{pmatrix} \mathbf{t} \\ \mathbf{u}\theta \end{pmatrix} \quad such\ that \quad [\mathbf{r}] \stackrel{def}{=} \begin{pmatrix} [\mathbf{u}]_\wedge & \mathbf{t} \\ \mathbf{0_{1\times 3}} & 0 \end{pmatrix} \stackrel{def}{=} logm\mathbf{T} \tag{11}$$

where $logm$ is the matrix logarithm and $\mathbf{T}$ is the 4×4 homogeneous matrix representation of $\mathbf{r}$.

Thus, the usual four steps of the Simplex S can be used:

$$reflection : \mathbf{r}_R = (1-\alpha)g + \alpha \mathbf{r}_W \tag{12}$$

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

$$expansion : \mathbf{r}_E = (1-\gamma)g + \gamma \mathbf{r}_R \tag{13}$$

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

$$contraction : \mathbf{r}_C = (1-\beta)g + \beta \mathbf{r}_W \tag{14}$$

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

$$\begin{aligned} shrinkage : \mathbf{r}'_G &= (\mathbf{r}_G + \mathbf{r}_B)/2 \\ \mathbf{r}'_W &= (\mathbf{r}_W + \mathbf{r}_B)/2 \end{aligned} \tag{15}$$

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 ε is a predefined eligible small distance, $val(S)$ is defined as

$$val(S) = \max\left(dist(\mathbf{r}_W,\ \mathbf{r}_B),\ dist(\mathbf{r}_W,\ \mathbf{r}_G),\ dist(\mathbf{r}_G,\ \mathbf{r}_B)\right) \tag{16}$$

and $dist$ is the distance between two vertices. By convention, the vertices are ordered as

$$f(\mathbf{r}_1) \leq f(\mathbf{r}_2) \leq \cdots \leq f(\mathbf{r}_{k+1}) \tag{17}$$

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

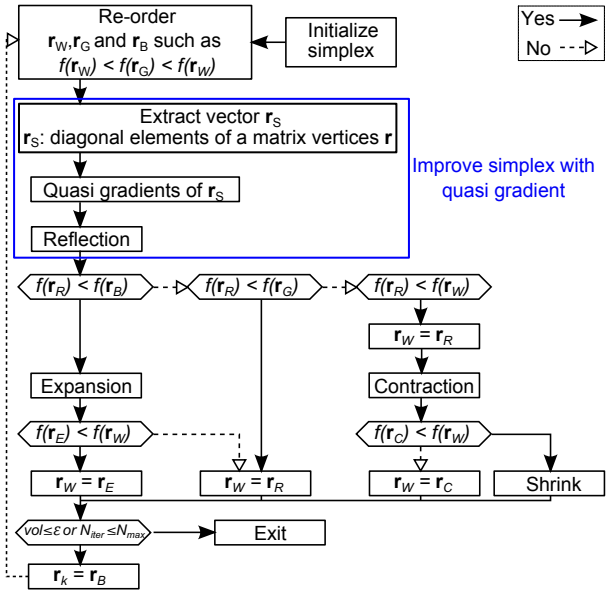

Fig. 5. Modified Simplex Algorithm.

The minimization of the cost-function using the Simplex algorithm is shown in Fig. 5. In our case, the Simplex was modified, by introducing the quasi-gradient convergence instead of reflection stage method [14], 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 non-linear method with 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, (12) is replaced with

$$\mathbf{r}_R = \mathbf{r}_B - \alpha \mathbf{Q} \tag{18}$$

where $\mathbf{Q}$ is the quasi-gradient vector based on the diagonal elements of the vertices matrix [14].

5 Registration *vs.* Visual Servoing

5.1 Image Transformation

First, the considered registration is defined as a rigid transformation between two images. Let us assume the transformation $\widehat{\mathbf{r}} \in SE(3) = \mathcal{R}(3) \times SO(3)$ 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 $\mathbf{MI}(\mathbf{I}_{b1}, \mathbf{I}_{b2})$:

$$\widehat{\mathbf{r}} = arg\min -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \;\mid\; \mathbf{r} \in SE(3) \tag{19}$$

where $\mathbf{r}$ is a possible rigid transformation.

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 $\widehat{\mathbf{r}}$ 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.

5.2 Visual Servoing

Let us assume that we have the cost-function shown in Fig. 3, then our objective is to find the global minimum

$$\widehat{\mathbf{r}} = arg \min_{\mathbf{r} \in SE(3)} -\mathbf{MI}\left[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}\right] \tag{20}$$

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 $\widehat{\mathbf{r}}^{-1}$ to the robot and start again (Fig. 7). 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 $\mathbf{r}_\mathbf{B}^\mathbf{n}$, the best vertex so far, which is associated to the best transformation ${}^\mathbf{0}\mathbf{T}_\mathbf{n} = e^{[\mathbf{r}_\mathbf{B}^\mathbf{n}]}$, with $[\mathbf{r}_\mathbf{B}^\mathbf{n}] = \begin{pmatrix} [\mathbf{u_n}\theta_\mathbf{n}]_\wedge & \mathbf{t_n} \\ 0 & 0 \end{pmatrix}$, from the initial to the current pose thanks

to the exponential mapping. Thus, applying directly the Simplex would require displacing the robot by

$$^{\mathbf{n-1}}\mathbf{T_n} = \left(^{\mathbf{0}}\mathbf{T_{n-1}}\right)^{-1}\ ^{\mathbf{0}}\mathbf{T_n} \tag{21}$$

where $^{\mathbf{0}}\mathbf{T_{n-1}} = e^{\begin{pmatrix}[\mathbf{u_{n-1}}\theta_{\mathbf{n-1}}]_\wedge & \mathbf{t_{n-1}} \\ 0 & 0\end{pmatrix}}$

This displacement will not be applied to the complete transformation $^{\mathbf{n-1}}\mathbf{T_n}$ found, because that may have the robot to take too large motion. Instead, we extract the screw $(\mathbf{\Delta t}, \mathbf{u}\theta)^\top$ associated to $^{\mathbf{n-1}}\mathbf{T_n}$ and convert it to a damped velocity over the sample period T_s which is $\mathbf{v} = \left(\lambda.\mathbf{\Delta t}\right)/T_s$ and $\omega = \left(\lambda.\mathbf{u\Delta}\theta\right)/T_s$.

Applying this velocity to the robot requires to update the Simplex vertex $\mathbf{r_B^n}$ according to the current (estimated) transformation (Fig. 6):

$$\left(\mathbf{r_B^n}\right)^{update} \Leftrightarrow {}^{\mathbf{0}}\mathbf{T_n}^{update} = \left(^{\mathbf{0}}\mathbf{T_{n-1}}\right)^{-1} e^{\begin{pmatrix}[\omega]_\wedge & \mathbf{v} \\ \mathbf{0} & 0\end{pmatrix}T_s} \tag{22}$$

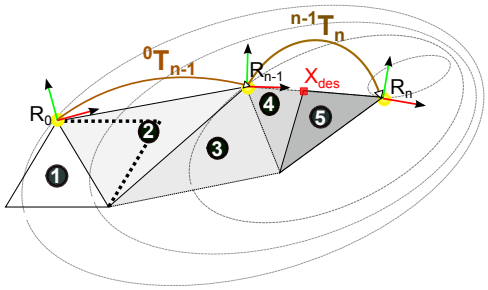

Fig. 6. Possible evolution of the Simplex.

6 Real-World Validation

6.1 Planar Positioning

Numerical Registration

The proposed numerical registration method is validated using two vocal folds images: real fluorescence and white light. These images taken from [1] were acquired in two different points of view with known pose as shown in Fig. 8. It can be highlighted that $\widehat{\mathbf{r}}$ between $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$ includes four parameters (x, y, θ and $zoom$). To be more realistic 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. 8). Then

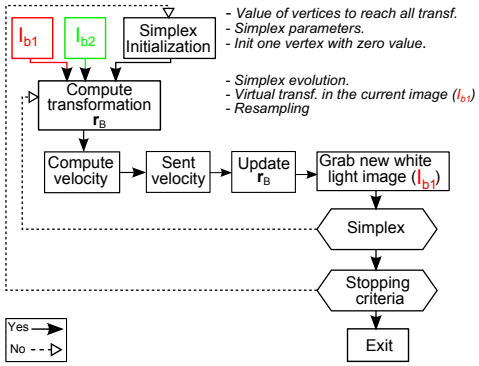

Fig. 7. MI-based visual servoing scheme.

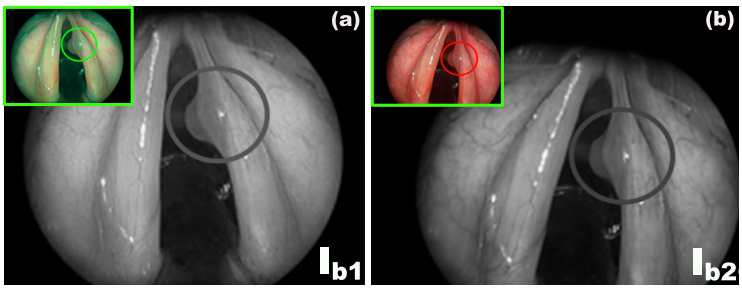

Fig. 8. (a) fluorescence image $\mathbf{I}_{b2}$ and (b) white light image $\mathbf{I}_{b1}$.

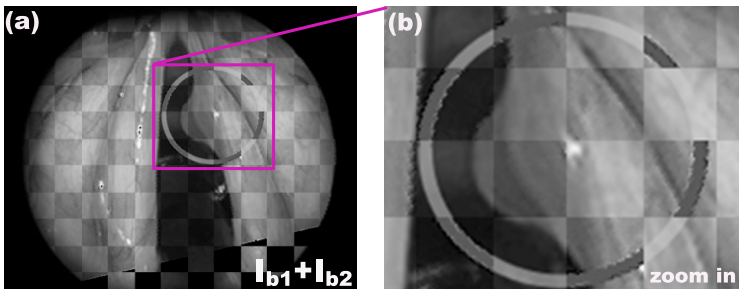

Fig. 9. Numerical registration results: (a) shows $\mathbf{I}_{b1}$ integrated in $\mathbf{I}_{b2}$, and (b) a zoom in the region of interest.

by analyzing Fig. 9(a), can be underlined the continuity of the combination $(\mathbf{I}_{b1} + \mathbf{I}_{b2})$, which relates to the high accuracy of the registration method. This accuracy is clearly visible on the zoom in the incision mark (Fig. 9(b)). For this example, the numerical values are summarized in Table 1.

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

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

Visual Servoing

For ethical reasons, we have not yet performed trials in a clinical set-up. Therefore, we validated the method on two benchmarks. The first one is a 3 DOF (x, y, θ) microrobotic cell (Fig. 10).

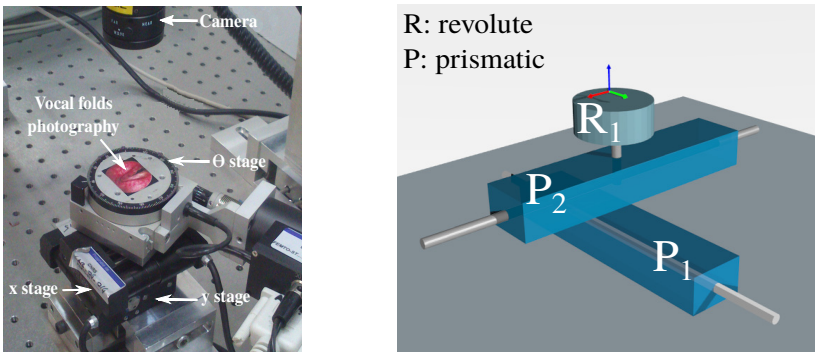

Fig. 10. Global view on the 3DOF experimental platform.

Firstly, the MI-based visual servoing is validated on monomodal images in aim to verify the validity of our controller. Figure 11(A) represents an example of white light images registration in visual servoing mode. More precisely, Fig. 11(A-a) and (A-b) represent the initial and desired images, respectively. In the same way, Fig. 11(A-c) and (A-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 good accuracy of our method.

Figure 11(C) shows the evolution of the velocities v_x, v_y and ω_z in the different DOF *versus* number of iterations. It can be underlined that the developed 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.

Secondly, vocal folds multimodal images are also used to test the proposed controller. In this scenario, the desired image is in fluorescence mode (prerecorded image) and the current images are in white light mode as it would be in

the surgical context. Figure 11(B-a) and (B-b) show the initial image $\mathbf{I}_{b1}$ and the desired image $\mathbf{I}_{b2}$, respectively. Figure 11(B-c) and (B-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. 11(D) shows the evolution of the velocities v_x, v_y and ω_z with respect number of iterations. It can be also underlined that the controller converges with the accuracy to the desired position despite the large difference between $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$.

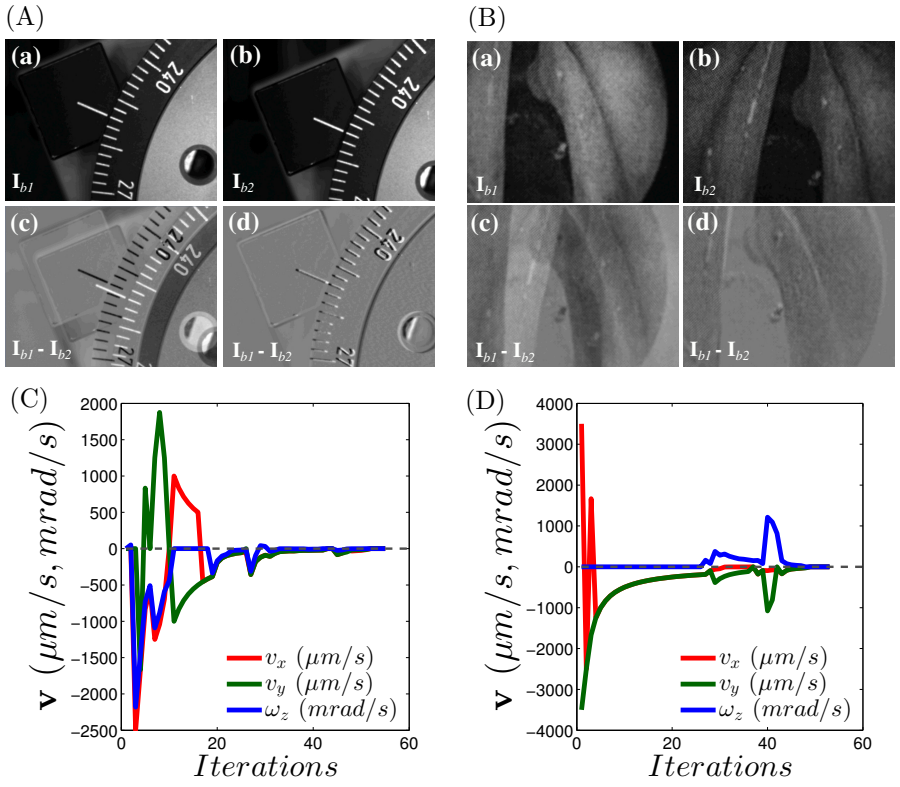

Fig. 11. Image snapshots acquired during the SE(2) positioning: (A) white light *vs.* white light images, (B) white light *vs.* fluorescence images. Velocities v_x, v_y and ω_z (in μm/s,mrad/s) *vs.* iterations in the case of: (C) white light *vs.* white light image (D) fluorescence *vs.* white light image.

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.

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

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

6.2 3D Positioning

Numerical Registration

This numerical registration was tested in the same condition as in the planar numerical registration experiment. However, in this case the transformation between $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$ is $\widehat{\mathbf{r}} \in SE(3)$. As in the previous experiment, we use the fluorescence image (Fig. 12(a)) *vs.* white light (Fig. 12(a)) image, with circular trajectory of the laser spot draw by the surgeon in both images. The initial Cartesian error between the desired image $\mathbf{I}_{b1}$ and the current image $\mathbf{I}_{b2}$, was $\mathbf{r}$ = (30 mm, 30mm, 40mm, 4°, 10°, 5°).

Again in this experiment we can see overlapping between the reference and the transformed image in the combined image Fig. 13(c). The resulting image is the sum between a region of current image (Fig. 13(a)) and the transformed one with the returned registration values (Fig. 13(b)) to show the continuity of the vocal fold shape. Besides, the real final error is $\delta\mathbf{r}$ = (0.22mm, 1.29mm, 9.5mm, 0.29°, 0.86°, 1.02°), with a computation time of 6.564 seconds.

Visual Servoing

The previous experiment on the visual servoing was extended to the 6 DOF

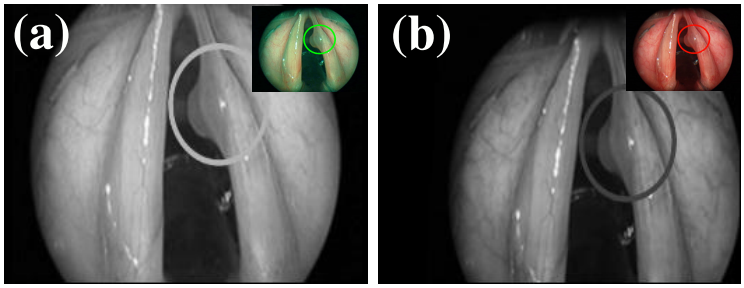

Fig. 12. (a) fluorescence image $\mathbf{I}_{b2}$ and (b) white light image $\mathbf{I}_{b1}$.

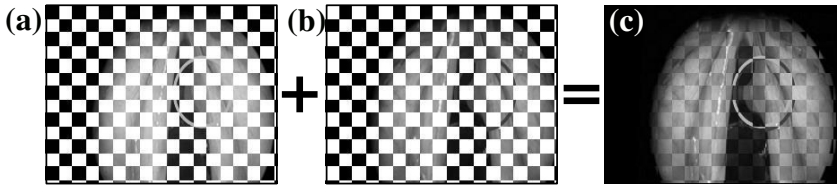

Fig. 13. Numerical registration results: (a) shows a sample region of $\mathbf{I}_{b1}$, (b) shows a sample region of $\mathbf{I}_{b2}$ after applying the numerical registration transformation, and (c) the combination of the images (a)+(b).

robot platform with an eye-to-hand configuration as shown in the Fig. 14(left). The test consists in the validation of our controller without any information of the setup as an interaction matrix or calibration parameters.

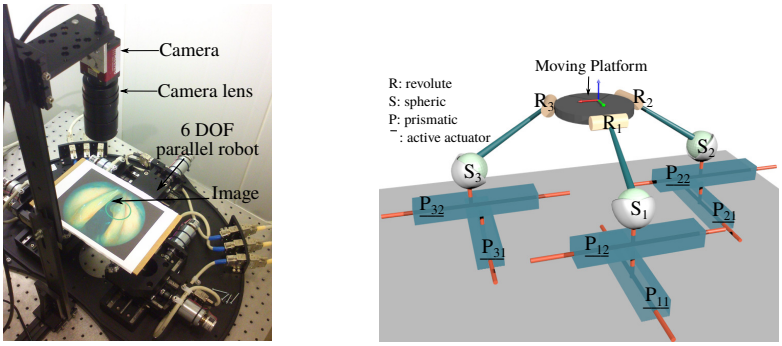

Fig. 14. Global view on the 6 DOF experimental platform.

The approach consists of 3D positioning of the robot based on desired image Fig. 15(a) (planer image (i.e., photography of vocal fold)) from current image Fig. 15(b) chosen arbitrary at the workspace of the robot. To do so, the robot

is placed at an initial position $\mathbf{r}$ = (-6mm, 6mm, 75mm, -1° , -1°, -1°) and must reach the desired position $\mathbf{r}^*$ = (6mm, -6mm, 74mm, -4°, 2°, 1°). While, the Fig. 15(c) presents the initial image difference ($\mathbf{I_{b1}} - \mathbf{I_{b2}}$) and Fig. 15(d) the final image difference when the controller reaches the desired position. The positioning errors in each DOF are computed using the robot encoders. The final error obtained is $\delta\mathbf{r}$ = (1.22mm, 0.352mm,0.320mm, 1.230°, 1.123°, 0.623°). By analyzing this numerical value, it can be underlined the convergence of the proposed method.

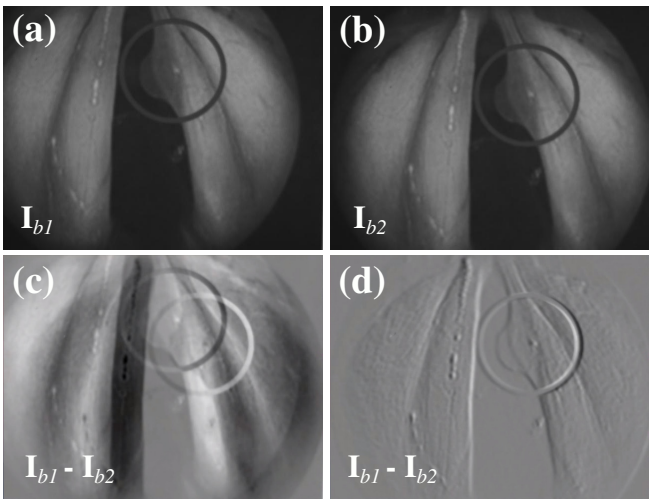

Fig. 15. Image sequence captured during the positioning task. (a) desired image $\mathbf{I_{b1}}$, (b) current image $\mathbf{I_{b2}}$, (c) initial difference $\mathbf{I_{b1}} - \mathbf{I_{b2}}$ and (d) final difference $\mathbf{I_{b1}} - \mathbf{I_{b2}}$ showing that the controller reaches the desired position.

In Fig. 16(a)-(b) illustrate the velocities $\mathbf{v}$ evolution sends to the robot during the positioning task relative to the number of iterations (each iteration takes 0.5 seconds). Furthermore, the mutual information values evolution decay is presented in Fig. 16(c) with respect to the number of iterations.

7 Conclusion

In this paper, a novel metric visual servoing-based on mutual information has been presented. Unlike the traditional methods, the developed approach was based only on the use of a modified Simplex optimization. It has been shown that the designed controller works even in the presence of many local minima in the mutual information cost-function. Beside this, the controller has shown good behavior in terms of repeatability and convergence. Also, we have validated the controller in SE(3) using a 6 DOF robot.

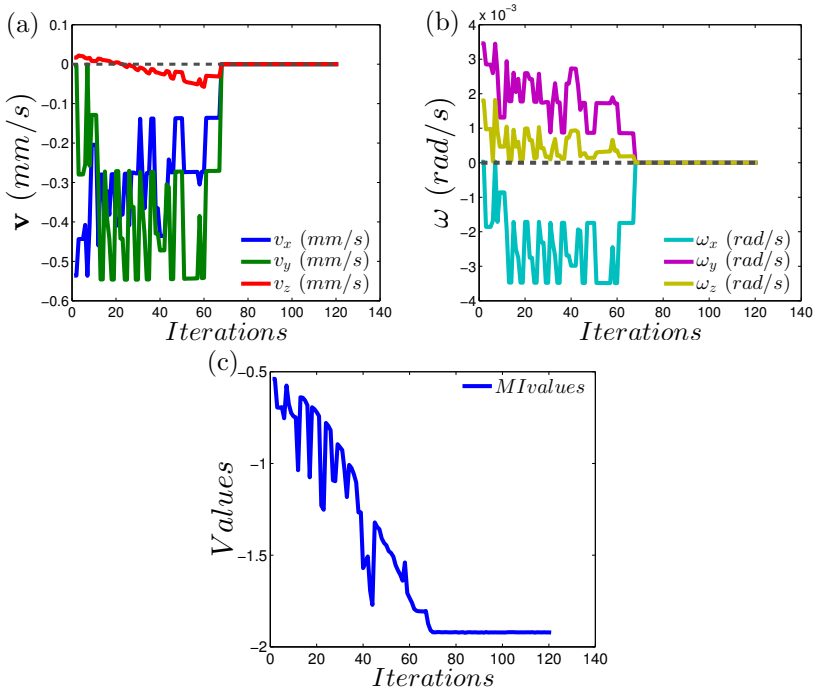

Fig. 16. (a) translation velocities $\mathbf{v}$ (in mm/s), (b) rotation velocities ω (in rad/s), (c) mutual information values evolution.

Future work will be devoted to optimize the computation time to reach the video rate and improve the velocity control trajectories.

Acknowledgements

This work was supported by μRALP, the EC FP7 ICT Collaborative Project no. 288663 [3], by French ANR NEMRO no ANR-14-CE17-0013-001, and by LABEX ACTION, the French ANR Labex no. ANR-11-LABX-0001-01 [4].

References

1. Arens, C., Dreyer, T., Glanz, H. Malzahn, K.: Indirect autofluorescence laryngoscopy in the diagnosis of laryngeal cancer and its precursor lesions. European Archives of Oto-Rhino-Laryngology and Head & Neck vol. 261(2), pp. 71–76, (2004)
2. Chaumette, F., Hutchinson, S.: Visual servo control, part 1 : Basic approaches. IEEE Robotics and Automation Magazine. vol. 13(1), pp. 82–90, (2006)

[3] `http://www.microralp.eu`
[4] `http://www.labex-action.fr`

3. Collewet, C., Marchand, E.: Photometric visual servoing. IEEE Trans. on Robotics. vol. 27(4), pp. 828–834 (2011)
4. Dame, A., Marchand, E.: Entropy-based visual servoing. In: IEEE Int. Conf. on Robotics and Automation. pp. 707–713, (2009)
5. Dame, A., Marchand, E.: Mutual information-based visual servoing. IEEE Trans. on Robotics. vol. 27(5), pp. 958–969, (2011)
6. Degani, A., Choset, H., Wolf, A., Zenati, M.A.: Highly articulated robotic probe for minimally invasive surgery. IEEE International Conference on Robotics and Automation. pp. 4167–4172. (2006)
7. Dowson, N., Bowden, R.: A unifying framework for mutual information methods for use in non-linear optimisation. Lecture Notes in Computer Science, vol. 3951, pp. 365–378, (2006)
8. Eckel, H., Berendes, S., Damm, M., Klusmann, J.: Suspension laryngoscopy for endotracheal stenting. Laryngoscope. vol. 113, pp. 11–15 (2003)
9. Jackel, M., Martin, A., Steine, W.: Twenty-five years experience with laser surgery for head and neck tumors. European Archives of Oto-Rhino-Laryngology. vol. 264, pp. 577–585, (2013)
10. Kelley, C.: Iterative Methods for Optimization. Frontiers in Applied Mathematics, vol. 18, (1999)
11. Marchand, E., Collewet, C.: Using image gradient as a visual feature for visual servoing. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. pp. 5687–5692, (2010)
12. Miura, K., Hashimoto, K., Gangloff, J., de Mathelin, M.: Visual servoing without jacobian using modified simplex optimization. IEEE International Conference on Robotics and Automation on. pp. 3504–3509,(2005)
13. Nelder, A., Mead, R.: A simplex method for function minimization. Computer-Journal. vol. 7, pp. 308–313, (1965)
14. Pham, N., Wilamowskial, B.: Improved nelder mead's simplex method and applications. Journal of computing. vol. 3(3), pp. 55–63, (2011)
15. Sevick-Muraca, E.: Fluorescence-enhanced optical imaging and tomography for cancer diagnostics. In: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on. vol. 2, pp. 1482–1485 (2004)
16. Tamadazte, B., Andreff, N.: Weakly calibrated stereoscopic visual servoing for laser steering: Application to phonomicrosurgery. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. pp. 743–748, (2014)
17. Upile, T., Jerjes, W., Sterenborg, H.J., El-Naggar, A.K., Sandison, A., Witjes, M.J., Biel, M.A., Bigio, I., Wong, B.J., Gillenwater, A., et al.: Head & neck oncology. Head & Neck. vol. 1(1), pp. 16, (2009)
18. Zitová, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing. vol. 21(11), pp. 977–1000, (2003)
19. Ourak, M., Tamadazte, B., Andreff, N., Marchand, E.: Visual servoing-based registration of multimodal images. International Conference on Informatics in Control, Automation and Robotics. pp. 1–8, (2015)