



HAL
open science

Multimodal Image Registration and Visual Servoing

Mouloud Ourak, Brahim Tamadazte, Nicolas Andreff, Eric Marchand

► **To cite this version:**

Mouloud Ourak, Brahim Tamadazte, Nicolas Andreff, Eric Marchand. Multimodal Image Registration and Visual Servoing. Lecture Notes in Electrical Engineering, 383, Springer, pp.157 - 175, 2016, 10.1007/978-3-319-31898-1_9 . hal-01433996

HAL Id: hal-01433996

<https://inria.hal.science/hal-01433996>

Submitted on 13 Jan 2017

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.

Multimodal Image Registration and Visual Servoing

M. Ourak, B. Tamadazte, N. Andreff and E. Marchand

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

11 1 Introduction

12 This work is grounded into robot assisted laser phonosurgery (RALP). The current
 13 gold standard procedure for the vocal folds surgery is certainly suspension micro-
 14 laryngoscopy (Fig. 1a) which requires direct visualization of the larynx and the
 15 trachea as proposed in [9]. This system is widely deployed in hospitals but it features
 16 many drawbacks related to patient and staff safety and comfort. Therefore, alterna-
 17 tive endoscopic approaches are under investigation: the extended use of the HARP
 18 (Highly Articulated Robotic Probe) highly flexible robot, designed for conventional
 19 surgery [6] or the use of an endoscopic laser micro-manipulator [17] (Fig. 1b). In
 all aforementioned cases, cancer diagnosis can be performed thanks to fluorescence

M. Ourak (✉) · B. Tamadazte · N. Andreff
 FEMTO-ST Institute, AS2M Department, Université Bourgogne
 Franche-Comté/CNRS/ENSMM, 24 Rue Savary, 25000 Besançon, France
 e-mail: mouloud.ourak@femto-st.fr

E. Marchand
 Université de Rennes 1, IRISA, Rennes, France

© Springer International Publishing Switzerland 2016

J. Filipe et al. (eds.), *Informatics in Control, Automation and Robotics 12th International Conference, ICINCO 2015 Colmar, France, July 21-23, 2015 Revised Selected Papers*, Lecture Notes in Electrical Engineering 383, DOI 10.1007/978-3-319-31898-1_9

1

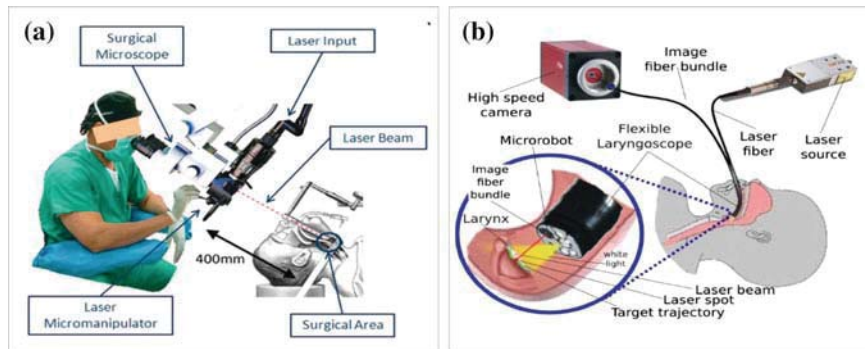


Fig. 1 Global view of the microphonosurgery system: **a** the current laser microphonosurgery system and **b** the targeted final system

20 imaging [16], (a few) days before the surgical intervention. The latter is usually
 21 performed under white-light conditions because fluorescence may require longer
 22 exposure time than real time can allow. Therefore, during a surgical intervention the
 23 fluorescence diagnosis image must be registered to the real-time white light images
 24 grabbed by the endoscopic system in order to define the incision path of the laser
 25 ablation or resection. Registration can be done either numerically or by physically
 26 servoing the endoscope to the place where the preoperative fluorescence image was
 27 grabbed.

28 In this paper, our aim is to control a robot based on direct visual servoing, i.e.
 29 using image information coming from white light and fluorescence sensors. Several
 30 visual servoing approaches based on the use of features (line, Region of interest
 31 (ROI)) [2] or the image global information (gradient [11], photometry [3] or mutual
 32 information [5]) can be used. Nevertheless, the use of mutual information (MI)
 33 in visual servoing problems has proved to be especially effective in the case of
 34 multimodal and less contrasted images [4]. In fact, these control techniques assume
 35 that the kinematic model of the robot and the camera intrinsic parameters are at
 36 least partially known, but would fail if the system parameters were fully unknown.
 37 In practice, the initial position cannot be very distant from the desired position to
 38 ensure convergence. To enlarge the stability domain, [12] proposed to use the Simplex
 39 method [13] instead of the usual gradient-like methods (which require at least a rough
 40 calibration of the camera and a computation of the camera/robot transformation).
 41 However, the work in [12] relies on the extraction from the image of geometrical
 42 visual features.

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

- 46 1. Calibration procedures are often difficult to perform, especially by non-
 47 specialist operators i.e., clinicians.

- 48 2. Surgeons entering in the operating room are perfectly sterilized to avoid any risk
49 of contamination, and then it is highly recommended to limit the manipulation of
50 the different devices inside the operating room.

51 For these reasons, we opted for uncalibrated and model-free multimodal registration
52 and visual servoing schemes using mutual information as a global visual feature and
53 a Simplex as optimization approach. Thereby, it is not necessary to compute the
54 interaction matrix (Jacobian image); the kinematic model of the robot may be totally
55 unknown, without any constraint in the initial position of the robot with respect to its
56 desired position. A preliminary version of this work was presented in [14] in the case
57 of planar positioning and is extended in this paper to positioning in the 3D space.

58 This paper is structured as follows: Sect. 2 explains the medical application of
59 the proposed approach. Section 3 gives the basic background on mutual information.
60 Section 4 presents a modified Simplex method. Section 5 describes multimodal reg-
61 istration and multimodal visual servoing. Finally, Sect. 6 deals with the validation
62 results.

63 2 Medical Application

64 The vocal folds are situated at the center and across the larynx and form a V-shaped
65 structure. They are used to create the phonation by modulating the air flow being
66 expelled from the lungs through quasi-periodic vibrations. They can be affected by
67 benign lesions, such as cysts or nodules (for instance, when they are highly stressed,
68 e.g. when singing) or, in the worst case, cancer tumors (especially for smokers). These
69 lesions change the configuration of the folds and thereby the patient's voice. Nowa-
70 days, medical tools can be used to suppress this trouble and recover the original voice
71 in particular for cyst and nodules. Appeared in 1960, *phonosurgery*—the surgery of
72 the vocal folds—can be divided into laryngoplastic, laryngeal injection, renovation
73 of the larynx and phonomicrosurgery. Specifically, laser phonomicrosurgery con-
74 sists of a straight rigid laryngoscope, a stereoscopic microscope, a laser source, and
75 a controlled 2DOF device to orient the laser beam [8], as shown in Fig. 1a. Nev-
76 ertheless, the current system requires extreme skill from the clinician. Specifically,
77 high dexterity is required because both the laser source is located out of the patient,
78 400 mm away from the vocal folds. This distance increases the risk of inaccuracy
79 when the laser cutting process is running. Moreover, the uncomfortable position of
80 the patient's neck in a straight position all along the operation can be traumatic. The
81 drawbacks of the conventional procedure are taken into account in the new set-up
82 developed within the European project μ RALP, which consists on embedding all the
83 elements (i.e., cameras, laser and guided mirror) inside an endoscope Fig. 1b. More
84 precisely, the endoscope is composed of white light, high speed camera imaging
85 the laser evolution with 3D feedback to the clinician. Additionally, a low framerate,
86 high sensitivity fluorescence imaging system is to be used preoperatively to detect
87 cancerous lesions.

88 The global approach is based on the use of 2 degrees of freedom (DOF) to guide the
89 laser along the trajectory drawn by the surgeon on a preoperative fluorescence image.
90 However, since the preoperative image is not necessarily taken by the same instru-
91 ment on the same location, this approach requires the preoperative fluorescence
92 image (where the surgeon decides the trajectory) and the white light image (where
93 the control of the robot is developed) to be registered. This can be done in two ways:
94 registration or servoing. Registration deals with the estimation of the transformation
95 between both images, which can then be used to morph the fluorescence image onto
96 the real-time endoscopic image flow (for instance, as an augmented reality). Visual
97 servoing deals with bringing the endoscope back to the place where the fluorescence
98 image was grabbed and stabilizing it in that configuration, which amounts to a phys-
99 ical registration and should turn useful in many other applications, such as surgery
100 in the stomach to compensate for physiological motions.

101 3 Mutual Information and Registration

102 In the literature, multimodal image registration has been widely discussed. Zitova
103 et al. [19] classified registration techniques for medical applications into two main
104 categories: area-based and features-based methods. In these cases, the registration
105 process follows mainly four steps: feature detection, feature matching, transforma-
106 tion estimation, and image resampling. As previously stated, our approach is based
107 on mutual information rather than geometrical visual features. Therefore, the most
108 critical steps (feature detection and matching) of a conventional registration method
109 are removed. Instead, from the joint and marginal entropy of two images, it is possi-
110 ble to compute their similarities. This means that the higher the mutual information
111 is, the better the images are aligned [4] (Fig. 2).

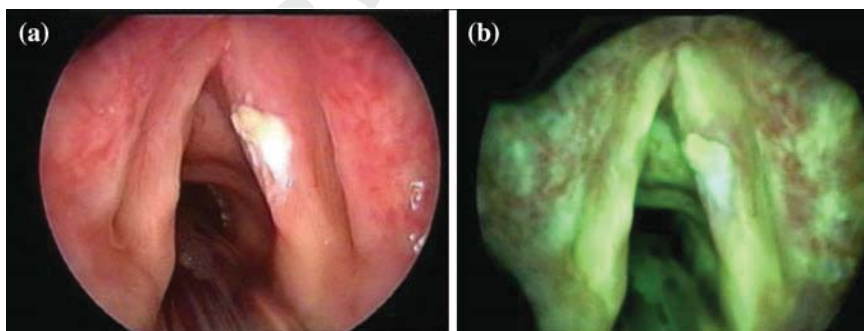


Fig. 2 Vocal folds endoscopic images: **a** white light endoscopic image, **b** fluorescence endoscopic image [18]

112 3.1 Mutual Information in the Image

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

$$115 \quad \mathbf{H}(\mathbf{I}) = - \sum_{i=0}^{N_I} p_{\mathbf{I}}(i) \log_2(p_{\mathbf{I}}(i)) \quad (1)$$

116 where $\mathbf{H}(\mathbf{I})$ represents the marginal entropy, also called Shannon entropy of an image
117 \mathbf{I} ; $i \in [0, N_I]$ (with $N_I = 255$) defines a possible gray value of an image pixel; and
118 $p_{\mathbf{I}}$ is the probability distribution function, also called marginal probability of i . This
119 can be estimated using the normalized histogram of \mathbf{I} .

120 Moreover, the entropy between two images \mathbf{I}_1 and \mathbf{I}_2 is known as joint entropy
121 $\mathbf{H}(\mathbf{I}_1, \mathbf{I}_2)$. It is defined as the joint variability of both images

$$122 \quad \mathbf{H}(\mathbf{I}_1, \mathbf{I}_2) = - \sum_{i=0}^{N_{I_1}} \sum_{j=0}^{N_{I_2}} p_{\mathbf{I}_1, \mathbf{I}_2}(i, j) \log_2(p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)) \quad (2)$$

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

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

$$129 \quad \mathbf{MI}(\mathbf{I}_1, \mathbf{I}_2) = \mathbf{H}(\mathbf{I}_1) + \mathbf{H}(\mathbf{I}_2) - \mathbf{H}(\mathbf{I}_1, \mathbf{I}_2) \quad (3)$$

130 and can be expressed using the marginal probability $p_{\mathbf{I}}$ and joint probability
131 $p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)$, by replacing (1) and (2) in (3) with some mathematical manipulations

$$132 \quad \mathbf{MI}(\mathbf{I}_1, \mathbf{I}_2) = \sum_{i,j} p_{\mathbf{I}_1, \mathbf{I}_2}(i, j) \log \left(\frac{p_{\mathbf{I}_1, \mathbf{I}_2}(i, j)}{p_{\mathbf{I}_1}(i) p_{\mathbf{I}_2}(j)} \right) \quad (4)$$

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

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

$$141 \quad \mathbf{I}_{b1}(k) = \mathbf{I}_1(k) \frac{N_c}{r_{max}} \quad \text{and} \quad \mathbf{I}_{b2}(k) = \mathbf{I}_2(k) \frac{N_c}{t_{max}} \quad (5)$$

142 where $t_{max} = r_{max} = 255$ and N_c are the new bin-size of the joint histogram and
 143 $\mathbf{I}_{b1}, \mathbf{I}_{b2}$ are the new gray level value of \mathbf{I}_1 and \mathbf{I}_2 , respectively.

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

$$151 \quad p_{\mathbf{I}_{b1}\mathbf{I}_{b2}}(i, j) = \frac{1}{N_k} \sum_k \psi(i - \mathbf{I}_{b1}(k)) \psi(j - \mathbf{I}_{b2}(k)) \quad (6)$$

$$152 \quad p_{\mathbf{I}_{b1}}(i) = \frac{1}{N_k} \sum_k \psi(i - \mathbf{I}_{b1}(k, x)) \quad (7)$$

$$153 \quad p_{\mathbf{I}_{b2}}(j) = \frac{1}{N_k} \sum_k \psi(j - \mathbf{I}_{b2}(k)) \quad (8)$$

154
 155 with N_k is the number of pixels in the new images \mathbf{I}_{b1} and \mathbf{I}_{b2} and ψ is the used
 156 B-spline function.

157 4 Simplex-Based Registration

158 This section deals with the method for solving the mutual information maximization
 159 problem. However, before describing the chosen optimization approach among the
 160 many existing ones [10] to solve this problem, it is necessary to know the exact shape
 161 of the cost-function in the case of bimodal images (fluorescence vs. white light) of
 162 the vocal cords.

163 In practice, rather than maximizing mutual information, we minimize the cost-
 164 function

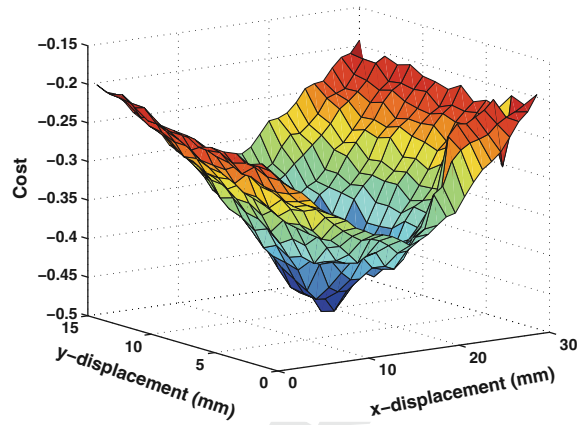
$$165 \quad \mathbf{f}(\mathbf{r}) = -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \quad (9)$$

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

$$168 \quad \hat{\mathbf{r}} = \arg \min_{\mathbf{r} \in SE(3)} \mathbf{f}(\mathbf{r}) \quad (10)$$

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

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



171 **4.1 Cost-Function Shape**

172 Figure 3 shows the computed cost-function in nominal conditions (i.e., the high def-
 173 inition images shown in Fig. 8). It has a global convex shape but still has many
 174 irregularities. Consequently, derivative based methods such as gradient descent could
 175 not necessarily guarantee convergence. Thereby, an unconstrained optimization tech-
 176 nique was chosen to overcome this problem, i.e., a modified Simplex algorithm.

177 **4.2 Modified Simplex Algorithm**

178 The Nelder-Mead Simplex algorithm [13] roughly works as follows. A Simplex
 179 shape S defined by vertices \mathbf{r}_1 to \mathbf{r}_{k+1} with $k = \dim(6)$ is iteratively updated until
 180 convergence using four operators: reflection, contraction, expansion, and shrinkage
 181 (see Fig. 4), defined on a linear space.

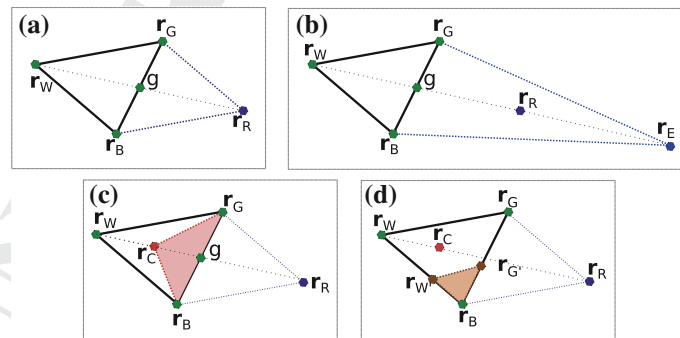


Fig. 4 Example of the Simplex steps: **a** reflection, **b** expansion, **c** contraction, and **d** shrinkage

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

$$184 \quad \mathbf{r} = \begin{pmatrix} \mathbf{t} \\ \mathbf{u}\theta \end{pmatrix} \quad \text{such that} \quad [\mathbf{r}] \stackrel{\text{def}}{=} \begin{pmatrix} [\mathbf{u}]_{\wedge} & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 0 \end{pmatrix} \stackrel{\text{def}}{=} \log m \mathbf{T} \quad (11)$$

185 where $\log m$ is the matrix logarithm and \mathbf{T} is the 4×4 homogeneous matrix repre-
186 sentation of \mathbf{r} .

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

$$188 \quad \text{reflection} : \mathbf{r}_R = (1 - \alpha)g + \alpha \mathbf{r}_W \quad (12)$$

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

$$191 \quad \text{expansion} : \mathbf{r}_E = (1 - \gamma)g + \gamma \mathbf{r}_R \quad (13)$$

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

$$193 \quad \text{contraction} : \mathbf{r}_C = (1 - \beta)g + \beta \mathbf{r}_W \quad (14)$$

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

$$195 \quad \begin{aligned} \text{shrinkage} : \mathbf{r}'_G &= (\mathbf{r}_G + \mathbf{r}_B)/2 \\ \mathbf{r}'_W &= (\mathbf{r}_W + \mathbf{r}_B)/2 \end{aligned} \quad (15)$$

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

197 Finally, the algorithm ends when $val(S) \leq \varepsilon$ where ε is a predefined eligible small
198 distance, $val(S)$ is defined as

$$199 \quad val(S) = \max (dist(\mathbf{r}_W, \mathbf{r}_B), dist(\mathbf{r}_W, \mathbf{r}_G), dist(\mathbf{r}_G, \mathbf{r}_B)) \quad (16)$$

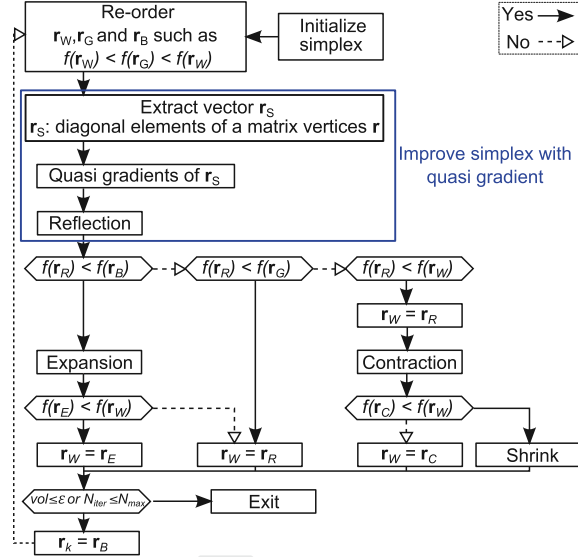
200 and $dist$ is the distance between two vertices. By convention, the vertices are ordered
201 as

$$202 \quad f(\mathbf{r}_1) \leq f(\mathbf{r}_2) \leq \dots \leq f(\mathbf{r}_{k+1}) \quad (17)$$

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

204 The minimization of the cost-function using the Simplex algorithm is shown in
205 Fig.5. In our case, the Simplex was modified, by introducing the quasi-gradient
206 convergence instead of reflection stage method [15], in order to improve the con-
207 vergence direction of f (without getting trapped in local minima) when the con-
208 troller approaches the desired position. This combination of an unconstrained and
209 non-linear method with a quasi-gradient technique allows a higher rate, faster and
210 smooth convergence speed. This returns to combine the advantages of a Simplex
211 and gradient-based optimization methods.

Fig. 5 Modified simplex algorithm



212 Therefore, (12) is replaced with

$$213 \quad \mathbf{r}_R = \mathbf{r}_B - \alpha \mathbf{Q} \quad (18)$$

214 where \mathbf{Q} is the quasi-gradient vector based on the diagonal elements of the vertices
215 matrix [15].

216 5 Registration Versus Visual Servoing

217 5.1 Image Transformation

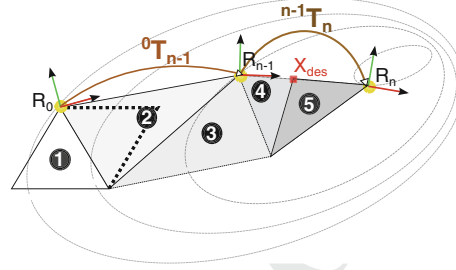
218 First, the considered registration is defined as a rigid transformation between two
219 images. Let us assume the transformation $\hat{\mathbf{r}} \in SE(3) = \mathcal{R}(3) \times SO(3)$ between the
220 white light image \mathbf{I}_{b1} and the fluorescence image \mathbf{I}_{b2} . Thereby, this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

$$222 \quad \hat{\mathbf{r}} = \arg \min -\mathbf{MI}[\mathbf{I}_{b1}(\mathbf{r}), \mathbf{I}_{b2}] \mid \mathbf{r} \in SE(3) \quad (19)$$

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

224 The process allowing to carry out this registration is operating as follows: acqui-
225 sition of both white light image \mathbf{I}_{b1} and fluorescence image \mathbf{I}_{b2} then computing
226 $\mathbf{MI}(\mathbf{I}_{b1}, \mathbf{I}_{b2})$. The obtained transformation $\hat{\mathbf{r}}$ from the first optimization is then applied

Fig. 6 Possible evolution of the simplex



227 to synthesize a new image $I_{b1}(\mathbf{r})$ from the image I_{b1} . These steps are repeated until
 228 the predefined stop criterion is reached (Fig. 6).

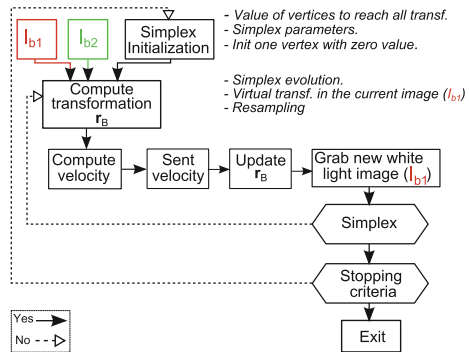
229 5.2 Visual Servoing

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

$$232 \quad \hat{\mathbf{r}} = \arg \min_{\mathbf{r} \in SE(3)} -\mathbf{MI} [I_{b1}(\mathbf{r}), I_{b2}] \quad (20)$$

233 A first way to move the robot so that the current (smoothed) image I_{b1} superimpose
 234 onto the desired fluorescence (smoothed) image I_{b2} is to use the look-than-move
 235 approach: let the Simplex method converge, then apply $\hat{\mathbf{r}}^{-1}$ to the robot and start
 236 again (Fig. 7). However, this requires a very fine tuning of the Simplex algorithm.
 237 The chosen approach allows interlacing the Simplex loop and the vision-based control
 238 loop. At each iteration n , the Simplex provides \mathbf{r}_B^n , the best vertex so far, which is
 239 associated to the best transformation ${}^0T_n = e^{[\mathbf{r}_B^n]}$, with $[\mathbf{r}_B^n] = \begin{pmatrix} [\mathbf{u}_n \theta_n]^\wedge & \mathbf{t}_n \\ 0 & 0 \end{pmatrix}$, from

Fig. 7 MI-based visual servoing scheme



240 the initial to the current pose thanks to the exponential mapping. Thus, applying
241 directly the Simplex would require displacing the robot by

$$242 \quad {}^{n-1}\mathbf{T}_n = ({}^0\mathbf{T}_{n-1})^{-1} {}^0\mathbf{T}_n \quad (21)$$

$$243 \quad \text{where } {}^0\mathbf{T}_{n-1} = e \begin{pmatrix} [{}^{\mathbf{u}_{n-1}}\theta_{n-1}]_{\wedge} & \mathbf{t}_{n-1} \\ 0 & 0 \end{pmatrix}$$

244 This displacement will not be applied to the complete transformation ${}^{n-1}\mathbf{T}_n$ found,
245 because that may have the robot to take too large motion. Instead, we extract the screw
246 $(\Delta\mathbf{t}, \mathbf{u}\theta)^\top$ associated to ${}^{n-1}\mathbf{T}_n$ and convert it to a damped velocity over the sample
247 period T_s which is $\mathbf{v} = (\lambda \cdot \Delta\mathbf{t})/T_s$ and $\omega = (\lambda \cdot \mathbf{u}\Delta\theta)/T_s$.

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

$$250 \quad (\mathbf{r}_B^n)^{update} \Leftrightarrow {}^0\mathbf{T}_n^{update} = ({}^0\mathbf{T}_{n-1})^{-1} e \begin{pmatrix} [\omega]_{\wedge} & \mathbf{v} \\ \mathbf{0} & 0 \end{pmatrix}_{T_s} \quad (22)$$

251 6 Real-World Validation

252 6.1 Planar Positioning

253 Numerical Registration

254 The proposed numerical registration method is validated using two vocal folds
255 images: real fluorescence and white light. These images taken from [1] were acquired
256 in two different points of view with known pose as shown in Fig. 8. It can be

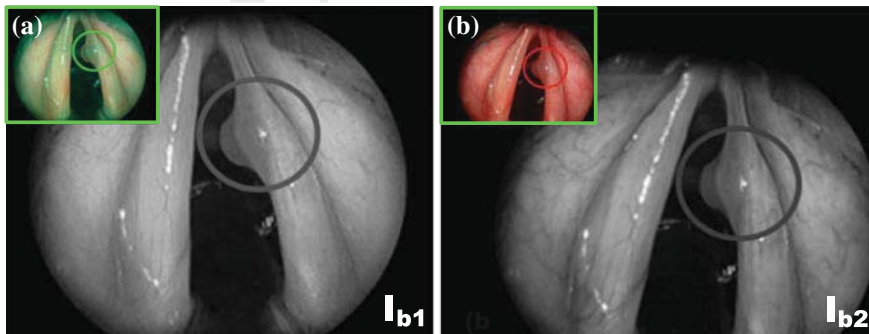


Fig. 8 a Fluorescence image I_{b2} and b white light image I_{b1}

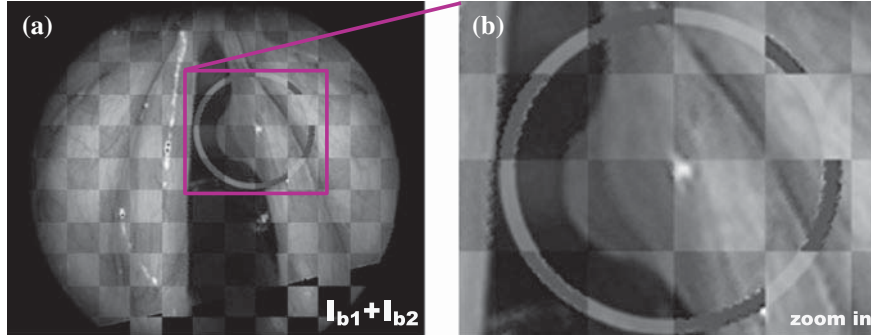


Fig. 9 Numerical registration results: **a** shows I_{b1} integrated in I_{b2} , and **b** a zoom in the region of interest

257 highlighted that $\hat{\mathbf{r}}$ between I_{b1} and I_{b2} includes four parameters (x , y , θ and $zoom$).
 258 To be more realistic in our validation tests, we added a circular trajectory (i.e., vir-
 259 tual incision mark done by a surgeon), to be tracked by the surgical laser spot, in the
 260 fluorescence image delimiting the tumor (Fig. 8). Then by analyzing Fig. 9a, can be
 261 underlined the continuity of the combination ($I_{b1} + I_{b2}$), which relates to the high
 262 accuracy of the registration method. This accuracy is clearly visible on the zoom in
 263 the incision mark (Fig. 9b). For this example, the numerical values are summarized
 264 in Table 1.

265 Visual Servoing

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

269 Firstly, the MI-based visual servoing is validated on monomodal images in aim
 270 to verify the validity of our controller. Figure 11a represents an example of white
 271 light images registration in visual servoing mode. More precisely, Fig. 11a(a, b)
 272 represent the initial and desired images, respectively. In the same way, Fig. 11a(c, d)
 273 show the initial and final error $I_{b1} - I_{b2}$. It can be noticed that the final position of
 274 the positioning platform matches perfectly with the desired position indicating good
 275 accuracy of our method.

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

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

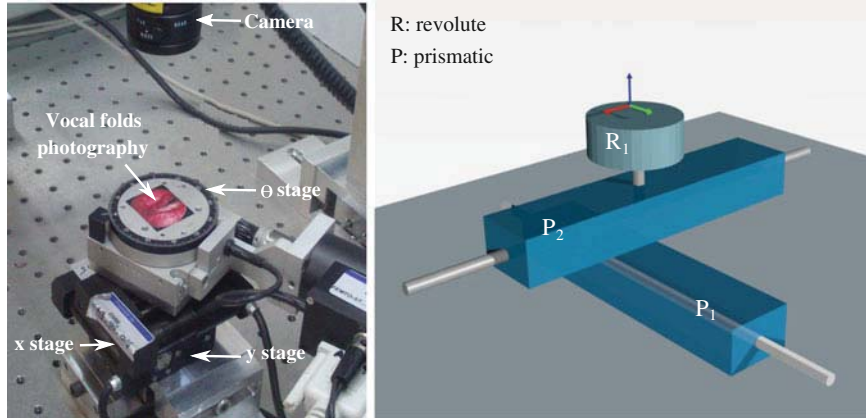


Fig. 10 Global view on the 3DOF experimental platform

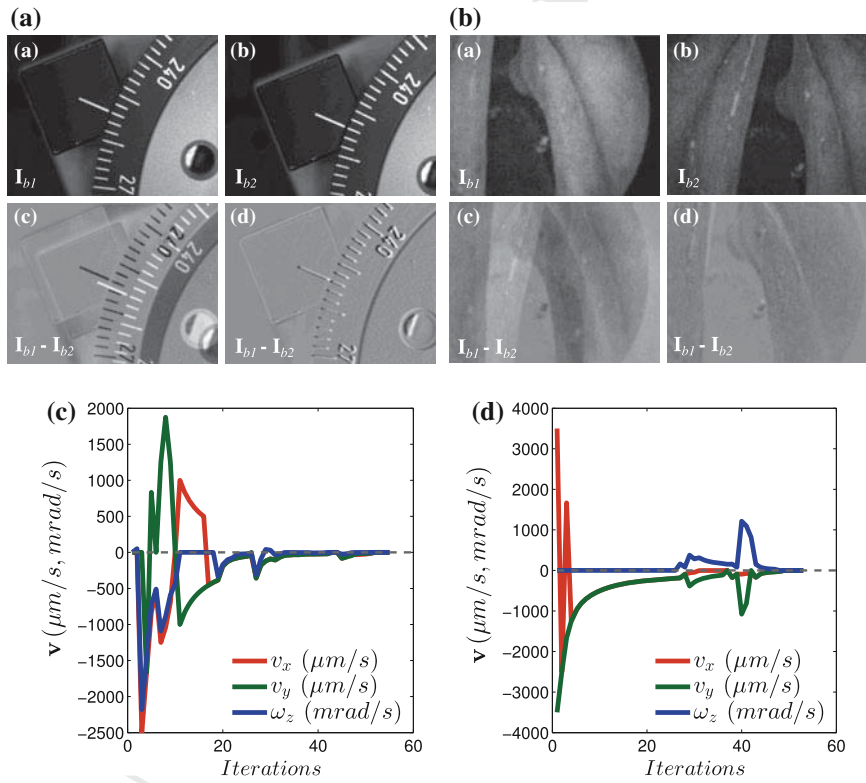


Fig. 11 Image snapshots acquired during the $SE(2)$ positioning: **a** white light versus white light images, **b** white light versus fluorescence images. Velocities v_x , v_y and ω_z (in $\mu\text{m/s}$, mrad/s) versus iterations in the case of: **c** white light versus white light image, **d** fluorescence versus white light image

276 Figure 11c shows the evolution of the velocities v_x , v_y and ω_z in the different
 277 DOF versus number of iterations. It can be underlined that the developed controller
 278 converges with accuracy in fifty iterations (*each iteration takes about 0.5 s*). Also,
 279 the speed varies in the iteration 40 because the Simplex after initialization found a
 280 new best minimum.

281 Secondly, vocal folds multimodal images are also used to test the proposed con-
 282 troller. In this scenario, the desired image is in fluorescence mode (pre-recorded
 283 image) and the current images are in white light mode as it would be in the surgical
 284 context. Figure 11b(a, b) show the initial image \mathbf{I}_{b1} and the desired image \mathbf{I}_{b2} , respec-
 285 tively. Figure 11b illustrate the error ($\mathbf{I}_{b1} - \mathbf{I}_{b2}$) during the visual servoing process.
 286 As shown in this figure, the controller converges also to the desired position with a
 287 good accuracy. Note that the image ($\mathbf{I}_{b1} - \mathbf{I}_{b2}$) is not completely gray (if two pixels
 288 are exactly the same, it is assigned the gray value of 128 for a better visualization of
 289 ($\mathbf{I}_{b1} - \mathbf{I}_{b2}$), this is due to the fact that both images are acquired from two different
 290 modalities, then the difference will never be zero (respectively 128 in our case).

291 In the same manner, Fig. 11d shows the evolution of the velocities v_x , v_y and ω_z
 292 with respect number of iterations. It can be also underlined that the controller con-
 293 verges with the accuracy to the desired position despite the large difference between
 294 \mathbf{I}_{b1} and \mathbf{I}_{b2} .

295 Additional validation tests were performed to assess the repeatability and behavior
 296 (convergence and robustness) of the controller. Therefore, for each test, the experi-
 297 mental conditions (lighting conditions, initial position and image quality) were delib-
 298 erately 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)

No.	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	

299 6.2 3D Positioning

300 Numerical Registration

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

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

313 Visual Servoing

314 The previous experiment on the visual servoing was extended to the 6 DOF robot
 315 platform with an eye-to-hand configuration as shown in the Fig. 14 (left). The test

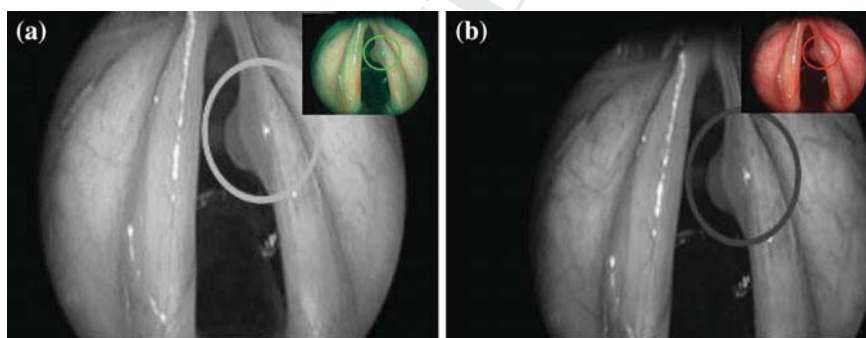


Fig. 12 a Fluorescence image \mathbf{I}_{b2} and b white light image \mathbf{I}_{b1}

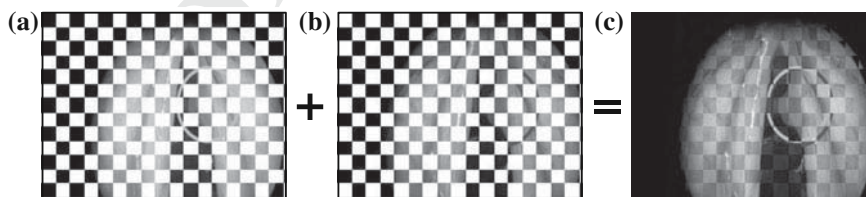


Fig. 13 Numerical registration results: a shows a sample region of \mathbf{I}_{b1} , b shows a sample region of \mathbf{I}_{b2} after applying the numerical registration transformation, and c the combination of the images (a) + (b)

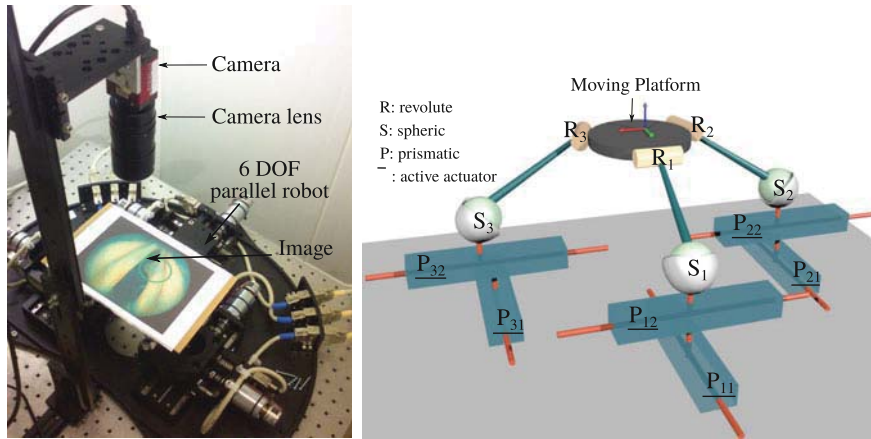


Fig. 14 Global view on the 6 DOF experimental platform

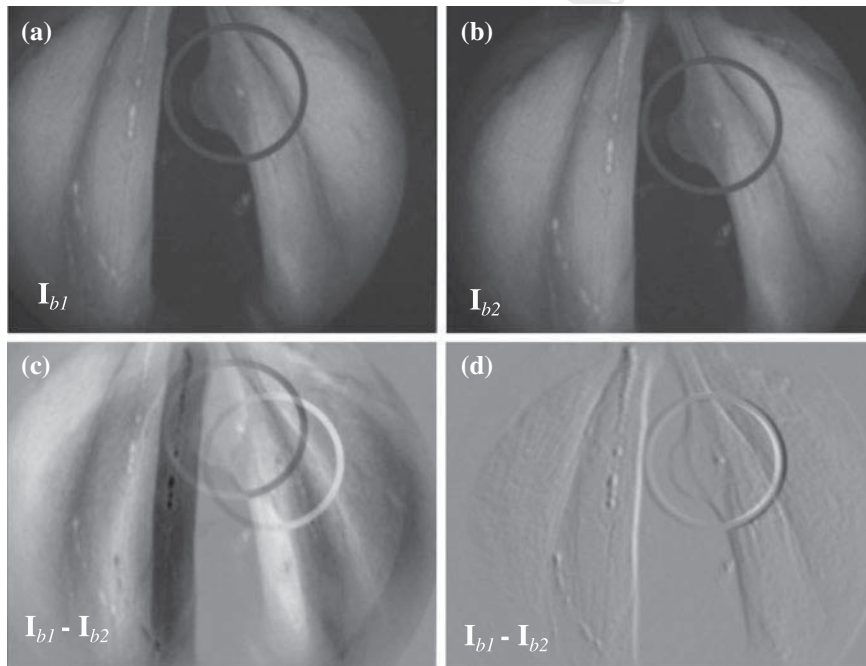


Fig. 15 Image sequence captured during the positioning task. **a** Desired image I_{b1} , **b** current image I_{b2} , **c** initial difference $I_{b1} - I_{b2}$ and **d** final difference $I_{b1} - I_{b2}$ showing that the controller reaches the desired position

316 consists in the validation of our controller without any information of the setup as
 317 an interaction matrix or calibration parameters.

318 The approach consists of 3D positioning of the robot based on desired image,
 319 Fig. 15a (planer image (i.e., photography of vocal fold)) from current image Fig. 15b
 320 chosen arbitrary at the workspace of the robot. To do so, the robot is placed at an initial
 321 position $\mathbf{r} = (-6, 6, 75 \text{ mm}, -1^\circ, -1^\circ, -1^\circ)$ and must reach the desired position $\mathbf{r}^* =$
 322 $(6, -6, 74 \text{ mm}, -4^\circ, 2^\circ, 1^\circ)$. While, the Fig. 15c presents the initial image difference
 323 $(\mathbf{I}_{b1} - \mathbf{I}_{b2})$ and Fig. 15d the final image difference when the controller reaches the
 324 desired position. The positioning errors in each DOF are computed using the robot
 325 encoders. The final error obtained is $\delta \mathbf{r} = (1.22, 0.352, 0.320 \text{ mm}, 1.230^\circ, 1.123^\circ,$
 326 $0.623^\circ)$. By analyzing this numerical value, it can be underlined the convergence of
 327 the proposed method.

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

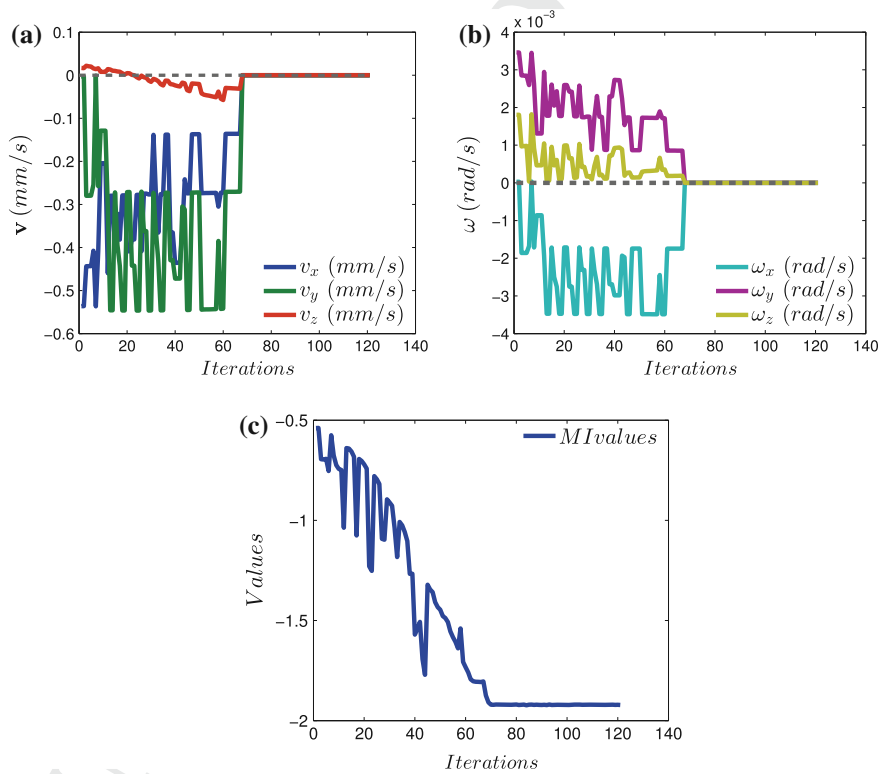


Fig. 16 a Translation velocities \mathbf{v} (in mm/s), b rotation velocities ω (in rad/s), c mutual information values evolution

332 **7 Conclusion**

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

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

342 **Acknowledgments** This work was supported by μ RALP, the EC FP7 ICT Collaborative Project
343 no. 288663 (<http://www.microralp.eu>), by French ANR NEMRO no ANR-14-CE17-0013-001, and
344 by LABEX ACTION, the French ANR Labex no. ANR-11-LABX-0001-01 ([http://www.labex-](http://www.labex-
345 action.fr)

346 **References**

- 347 1. Arens, C., Dreyer, T., Glanz, H., Malzahn, K.: Indirect autofluorescence laryngoscopy in the
348 diagnosis of laryngeal cancer and its precursor lesions. *Eur. Arch. Oto-Rhino-Laryngol. Head
349 Neck* **261**(2), 71–76 (2004)
- 350 2. Chaumette, F., Hutchinson, S.: Visual servo control, Part 1: basic approaches. *IEEE Robot.
351 Autom. Mag.* **13**(1), 82–90 (2006)
- 352 3. Collewet, C., Marchand, E.: Photometric visual servoing. *IEEE Trans. Robot.* **27**(4), 828–834
353 (2011)
- 354 4. Dame, A., Marchand, E.: Entropy-based visual servoing. In: *IEEE International Conference
355 on Robotics and Automation*, pp. 707–713 (2009)
- 356 5. Dame, A., Marchand, E.: Mutual information-based visual servoing. *IEEE Trans. Robot.* **27**(5),
357 958–969 (2011)
- 358 6. Degani, A., Choset, H., Wolf, A., Zenati, M.A.: Highly articulated robotic probe for minimally
359 invasive surgery. In: *IEEE International Conference on Robotics and Automation*, pp. 4167–
360 4172 (2006)
- 361 7. Dowson, N., Bowden, R.: A unifying framework for mutual information methods for use in
362 non-linear optimisation. *Lect. Notes Comput. Sci.* **3951**, 365–378 (2006)
- 363 8. Eckel, H., Berendes, S., Damm, M., Klusmann, J.: Suspension laryngoscopy for endotracheal
364 stenting. *Laryngoscope* **113**, 11–15 (2003)
- 365 9. Jackel, M., Martin, A., Steine, W.: Twenty-five years experience with laser surgery for head
366 and neck tumors. *Eur. Arch. Oto-Rhino-Laryngol.* **264**, 577–585 (2013)
- 367 10. Kelley, C.: *Iterative Methods for Optimization*. *Frontiers in Applied Mathematics*, vol. 18
368 (1999)
- 369 11. Marchand, E., Collewet, C.: Using image gradient as a visual feature for visual servoing. In:
370 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5687–5692 (2010)
- 371 12. Miura, K., Hashimoto, K., Gangloff, J., de Mathelin, M.: Visual servoing without Jacobian using
372 modified simplex optimization. In: *IEEE International Conference on Robotics and Automa-
373 tion*, pp. 3504–3509 (2005)
- 374 13. Nelder, A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**, 308–313
375 (1965)