

Positionnement 3-D d'un instrument flexible robotisé à l'aide d'une caméra monoculaire endoscopique : comparaison entre méthodes basées modèle et un apprentissage supervisé

Paolo Cabras¹ David Goyard¹ Florent Nageotte¹ Philippe Zanne¹ Christophe Doignon¹

¹ Laboratoire ICube (UMR CNRS 7357), Université de Strasbourg, France *

ICube - Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie
Télécom Physique Strasbourg 300 bd Sébastien Brant
CS 10413 - F-67412 Illkirch Cedex
cabras@unistra.fr

Résumé

La connaissance précise de la configuration spatiale des instruments flexibles actionnés est un facteur fondamental de réussite pour la chirurgie robotisée sans cicatrice. Pour ces opérations, les instruments de chirurgie, qui sont insérés dans les canaux d'un guide flexible, possèdent des capacités de mouvements propres qu'il faut contrôler.

Dans cet article, nous proposons d'estimer la position des instruments de la plateforme Anubis (Karl Storz) en utilisant exclusivement les images endoscopiques délivrées par une caméra miniature embarquée. Cette plateforme est une réalisation de guide flexible pour la chirurgie robotisée où les instruments insérés dans les canaux possèdent trois degrés de liberté (translation, rotation et déflexion). Avec la caméra endoscopique, les approches basées modèle ne délivrent pas toujours une précision suffisante à cause des incertitudes sur les paramètres. Pour surmonter cela, nous avons étendu le modèle pour autoriser une variation des paramètres et parallèlement nous avons exploré une technique d'apprentissage supervisé. A cette fin, des marqueurs ont été fixés sur les instruments et servent à entraîner un classifieur de type Adaboost sur les images in vivo segmentées. Pour effectuer la validation, les indices visuels sont utilisés également comme données d'entrée à une fonction préalablement apprise sur une vérité terrain fournie par un capteur magnétique. Finalement, une comparaison avec les méthodes basées modèle confirme la précision du positionnement 3-D obtenue en situation réelle.

Mots Clef

Estimation de pose, Robotique chirurgicale, Vision endoscopique, Apprentissage, Optimisation.

*This work was supported by French state funds managed by the ANR within the Investissements d'Avenir program (Labex CAMI) under reference ANR-11-LABX-0004. Authors thank Karl Storz company for providing the Anubis and the flexible instruments and IRCAD for access to the operating room.

Abstract

The spatial configuration of actuated flexible instruments is a fundamental key of success for applications in robotic no-scar surgery. During these operations, the instruments are inserted in the channels of a flexible guide equipped with a tiny endoscopic camera. In this paper we propose to estimate the position of the instruments of the Anubis platform (Karl Storz) using only the endoscopic images. For this mechanical system, flexible instruments have 3-DOF (translation and rotation in the channel and deflection). Application of standard approaches on this system do not provide good accuracy in any situations because of uncertainties on several model parameters. To cope with these uncertainties, the possibility to expand the model of the instrument and to apply a supervised learning method has been explored. With the help of visual markers attached to the instruments, an AdaBoost classifier has been manually trained on through in vivo images sequence. Subsequently, the resulting blobs are used as input data of an approximation function trained using ground truth information provided by a magnetic sensor. A comparison with model-based methods showed the potentialities of such a positioning approach on real situation.

Keywords

Pose estimation, Surgical robotics, Endoscopic vision, Learning-based method, Optimization.

1 Introduction

1.1 Introduction et motivations

Avec l'émergence des nouvelles techniques de chirurgie peu invasives telles que la chirurgie endoscopique transluminale par orifices naturels (NOTES - *Natural Orifices Transluminal Endoscopic Surgery*) [1], les systèmes mécaniques flexibles qui, par le passé, étaient employés à des fins de diagnostics ou pour réaliser des tâches mineures

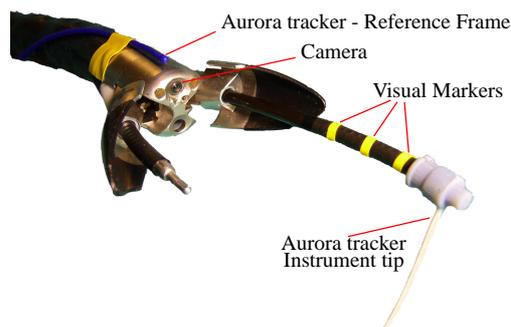


FIGURE 1 – Partie distale de la plateforme Anubis (Karl Storz) munie d'une caméra et d'instruments flexibles de chirurgie. Des marqueurs (en jaune) et un capteur magnétique (Aurora) ont été ajoutés pour mesurer le positionnement (voir section 4.2).

sont dorénavant pleinement utilisés lors des interventions chirurgicales pour la grande dextérité qu'ils procurent. La plateforme Anubis, développée par la société Karl Storz (Fig. 1) est un exemple d'un tel système pour l'assistance à des interventions chirurgicales à l'aide d'instruments flexibles. Vis-à-vis des techniques conventionnelles de chirurgie, celles qui utilisent des instruments de laparoscopie ou d'endoscopie flexibles nécessitent un haut degré d'abstraction. En effet, la coordination des gestes à réaliser avec les trajectoires effectives des instruments n'est pas facile à appréhender au travers des images endoscopiques, même stéréoscopiques. De plus, ces techniques requièrent la présence d'assistants pour maintenir la visibilité du siège de l'opération, prendre en compte les directives du chirurgien tout en manipulant le système durant toute la durée de l'opération. Ce sont deux des raisons qui ont permis aux systèmes d'assistance informatique et robotique d'être perçus comme des solutions intéressantes en fournissant une précision accrue et répétitive, amenant dans certains cas à réduire la durée totale de l'intervention.

Les systèmes robotiques flexibles conçus pour l'assistance à la chirurgie sont des systèmes pour lesquels il est difficile d'obtenir une précision suffisante (tout le long de la chaîne cinématique) lors de la commande des déplacements, et au regard de la technologie actuelle de l'actionnement (transmission par câbles, le plus souvent). En effet, la cinématique des systèmes flexibles est difficile à modéliser du fait des interactions mécaniques existantes entre les câbles et les gaines des portions flexibles, et des forces de friction. En conséquence, la position de l'organe terminal ne peut pas être estimée avec une précision suffisante à partir d'une mesure effectuée sur des codeurs proximaux.

1.2 Etat de l'art

Bien qu'il n'existe que très peu de travaux correspondant à la configuration particulière du système présenté dans [2], on note un nombre croissant de contributions concernant l'estimation de la pose de la partie distale des robots continus à sections multiples ainsi que sur la commande de tels robots. Si la modélisation géométrique et cinématique est

bien établie [3, 4], il n'en est pas de même pour la restitution réaliste de la dynamique. Des modèles qui prennent en compte la compensation des non-linéarités ont été proposés récemment [5], mais il ressort que le comportement dynamique réel de ces systèmes est difficile à établir.

Pour éviter cela, des asservissements ont été considérés, et réalisés grâce à l'ajout de capteurs électromagnétiques [6, 7]. Cependant il semble que ceux-ci ne soient pas très fiables dans un environnement *in vivo* car les mesures sont sensibles aux matériaux ferromagnétiques situés à proximité. L'idée d'effectuer un asservissement est tout de même celle qui prévaut car c'est le type d'approche qui peut compenser des incertitudes du modèle le long de la chaîne cinématique. En particulier, les asservissements visuels ont été explorés depuis plus d'une décennie sur les robots continus [8, 9, 10].

Ainsi, dans [8], un robot continu à sections multiples se déplaçant sur un plan perpendiculaire à l'axe de la caméra a été considéré. Les différentes *vertèbres*, séparant les sections successives, sont localisées grâce à la caméra, puis la courbure est calculée à l'aide d'une approximation circulaire des sections adjacentes.

L'hypothèse d'une transformation homographique entre une image de référence et l'image actuelle d'un robot continu ainsi que la connaissance du modèle direct ont été utilisés dans [9] pour estimer sa forme (et son évolution) à l'aide d'une caméra fixe. Pour ce faire des plans fictifs ont été calculés à partir de quadruplets de points stratégiquement répartis le long du corps du robot continu.

Plus récemment, une estimation de la configuration d'un robot continu à l'aide d'un réseau de caméras montées sur le robot a été proposée par Weber *et al.* [11]. Les auteurs présentent alors un algorithme qui consiste en une première étape de raffinement de l'estimation de la configuration, puis une étape d'ajustement de faisceaux, où coordonnées 3-D et paramètres de configuration du robot sont estimés simultanément.

En chirurgie mini-invasive, la vision endoscopique est de plus en plus employée pour localiser des instruments de laparoscopie [12]. Récemment, une méthode basée sur une classification supervisée a été présentée dans [13] pour détecter les pixels qui sont les probables d'appartenir aux instruments vis-à-vis des tissus vivants, à l'aide de classificateurs de type *random forest*. Ceux-ci servent par la suite à initialiser la forme apparente (projection d'un cylindre), et enfin à estimer la pose de l'instrument à l'aide d'un modèle 3-D et de la segmentation du contour de la silhouette de l'instrument à l'aide des fonctions de niveaux.

Deux approches pour estimer la pose d'un instrument flexible ont été proposées par Reilink *et al.* [14]. A l'aide d'un dispositif similaire à la plateforme Anubis, ces approches s'appuient sur des images endoscopiques et sur la connaissance de la configuration relative entre la caméra et l'instrument pour déterminer la position de l'extrémité de ce dernier, avec ou sans marqueurs artificiels. Comme dans les approches décrites précédemment sur la commande des

robots continus pour d'autres environnements, l'estimation de la pose 3-D nécessite de disposer d'un modèle précis de l'instrument robotisé. De tels modèles sont difficiles à obtenir et sont sujets à de fortes variabilités, dues par exemple à l'hystérésis et à la présence de jeu entre l'instrument et le canal opérateur.

D'autres travaux utilisent la stéréovision pour pallier le problème de l'estimation de la profondeur avec un capteur monoculaire [15, 16]. En particulier, Reiter *et al.* dans [16] propose une approche par apprentissage pour établir une relation entre les indices visuels dans les images endoscopiques et la position 3-D de l'instrument flexible. C'est donc assez naturellement que nous avons choisi d'explorer l'apport d'une technique d'apprentissage, mais avec seulement une seule caméra endoscopique.

Après une brève description du système expérimental, la section 3 décrit une analyse de l'approche basée modèle pour déterminer la position de l'extrémité de l'instrument flexible. Cette analyse servira par la suite à évaluer la nouvelle stratégie que nous proposons dans la section 4. Des résultats expérimentaux sont reportés et analysés puis une conclusion est donnée sur l'apport de notre approche.

2 Dispositif expérimental

Comme on peut le voir sur la Fig. 1, la partie distale de la plateforme Anubis est composée d'une caméra endoscopique et de deux instruments flexibles issus de deux canaux opérateurs inclus dans le corps de l'endoscope principal. Ces instruments possèdent trois degrés de liberté (ddl) commandables : une translation le long de l'axe du canal, une rotation autour de cet axe et une déflexion (représentés respectivement par les paramètres λ , ϕ et α sur la Fig. 2). Les environnements *in vivo* présentent souvent des images plus ou moins riches en texture, parfois avec de la fumée ou des traces de fluides. Ces variabilités perturbent l'acquisition et rendent l'interprétation des images et la segmentation délicates. Pour ces raisons, des marqueurs artificiels ont été ajoutés sur la partie flexible de l'instrument de chirurgie, sans que ceci perturbe sa fonction originale. En considérant par exemple trois marqueurs (Fig. 1 et Fig.

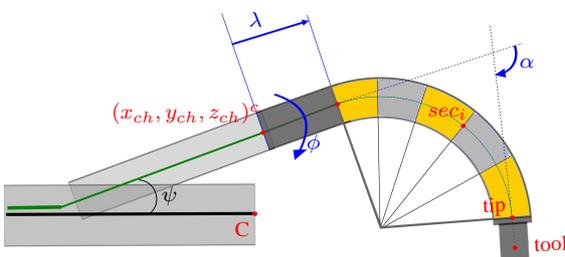


FIGURE 2 – Représentation de l'instrument flexible avec ses degrés de liberté (λ , ϕ , α) et les paramètres mécaniques intrinsèques : la position de la sortie du canal ($[x_{ch}, y_{ch}, z_{ch}]$) et l'orientation ψ de l'axe du canal par rapport à la caméra endoscopique, attachée au corps de l'endoscope principal.

2), la partie flexible est alors décomposée en 6 régions (ou sections) délimitées par les bords des marqueurs colorés et par l'instrument noir : la première région est sur la base de la partie flexible et la sixième région est à son extrémité distale. Ainsi, pour chaque région \mathcal{R}_i , deux points caractéristiques peuvent servir à définir des indices visuels le long du contour imagé (spots rouges sur la Fig. 4(a)), notés $\mathbf{p}_{i,u}$ pour le point supérieur et $\mathbf{p}_{i,l}$ pour le point inférieur (ceci n'est qu'une convention de dénomination puisque les points peuvent se retrouver dans un ordre opposé).

Afin de quantifier la précision des solutions basées sur la vision monoculaire, la mesure de position et d'orientation 3-D relative est également obtenue par un système de suivi placé sur l'instrument et sert de vérité terrain. Ce système est composé de deux capteurs électromagnétiques attachés à l'endoscope (Fig. 1), l'un est placé sur l'extrémité de la partie flexible, l'autre, utilisé comme repère de référence, est fixé sur l'endoscope principal, porteur de la caméra. Les mesures provenant de ces deux capteurs seront utilisées pour l'entraînement effectué par la méthode d'apprentissage (cf. section 4.2).

Ceci nécessite de mettre en oeuvre un étalonnage du dispositif (capteur électromagnétique vis-à-vis de la caméra) et d'étalonner la caméra par une méthode standard. A cette fin, un autre capteur électromagnétique est déplacé manuellement sur plusieurs positions connues d'un damier et chacune de ses positions est mesurée à la fois par le capteur électromagnétique embarqué sur l'endoscope et par la caméra endoscopique préalablement étalonnée (voir Fig. 3). La transformation euclidienne recherchée ${}^cT_{a1}$ entre le repère de la caméra et le repère de référence est estimée à l'aide des techniques standards de type *eye-to-hand* bien connues en robotique [17, 18]. Après reprojection des positions ponctuelles, nous obtenons une erreur RMS de [1.8, 1.3] pixels dans les images et de [0.5, 0.33, 0.16] mm dans l'espace 3-D pour les coordonnées x , y et z respectivement, au sein du volume utilisé.

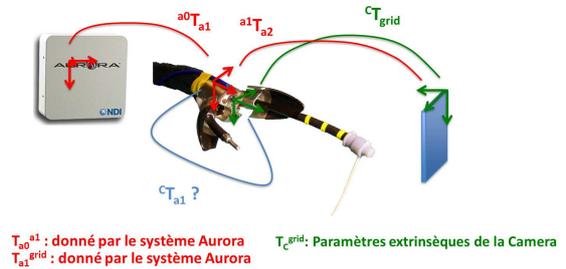


FIGURE 3 – Etalonnage de l'ensemble {capteur électromagnétique a_1 + caméra endoscopique C } à l'aide d'un damier et d'un second capteur électromagnétique mobile a_2 (base de référence électromagnétique a_0).

3 Pose 3-D basée sur un modèle

Dans cette partie, nous présentons la méthode de positionnement basée sur la connaissance du modèle cinématique de l'instrument flexible. Bien qu'il soit possible sous certaines conditions de résoudre ce problème analytiquement, les solutions les plus robustes sont le résultat d'un algorithme itératif et s'appuient sur les asservissements visuels virtuels [19, 12]. Ainsi, une image synthétique de la scène est calculée pour une configuration donnée de l'instrument et est comparée à l'acquisition vidéo qui est la référence. Des descripteurs communs aux deux images, tels que des points géométriques (calculés) / indices visuels (mesurés) servent à cette comparaison sous la forme d'une fonction de tâche, qui prend la forme d'une différence $\mathbf{e}^{(t)}$ entre l'état actuel (position des points dans l'image calculée) et l'état de référence (position des indices visuels dans l'image vidéo acquise). Puisqu'il n'y a pas d'acquisition en continu ni de déplacements réels mais une mise à jour de l'image calculée (t est plutôt une valeur d'itération), le terme de fonction de coût est préférable pour ce processus d'optimisation qui cherche à minimiser $\mathbf{e}^{(t)} = \mathbf{p} - \hat{\mathbf{p}}^{(t)}$ où \mathbf{p} est le vecteur des coordonnées des $2n$ points dans l'image réelle et $\hat{\mathbf{p}}^{(t)}$ celui des coordonnées des $2n$ projections calculées pour l'image virtuelle.

Par la suite les notations suivantes sont utilisées et correspondent aux coordonnées inhomogènes : \mathbf{P}_i est un point 3-D de la région \mathcal{R}_i ; $\mathbf{P}_{i,u}$ et $\mathbf{P}_{i,l}$ sont les correspondants 3-D des points 2-D supérieurs et inférieurs dans l'image de l'instrument délimitant la région \mathcal{R}_i , où $i = 1$ pour la base et $i = n = 6$ pour la dernière région englobant l'extrémité de la partie flexible. En rassemblant ces points, on note $\mathbf{P} = [\mathbf{P}_{1,u} \cdots \mathbf{P}_{6,u}, \mathbf{P}_{1,l} \cdots \mathbf{P}_{6,l}]^T$. Les mêmes notations sont prises avec une lettre minuscule, \mathbf{p} , pour représenter les points dans les images. Enfin le symbole \hat{g} signifie que la grandeur g est estimée.

3.1 Pose de l'instrument avec paramètres mécaniques constants

Dans cette section, la technique de Reilink est présentée [20]. Celle-ci suppose que le modèle cinématique de l'ensemble {endoscope + instrument flexible} est connu. A l'instar de la modélisation des robots rigides, le vecteur $\mathbf{q} = [\lambda, \phi, \alpha]^T$ rassemble les *variables articulaires*, degrés de liberté (ddl) de l'instrument, et il peut être établi une relation cinématique entre les variations des coordonnées d'un point 3-D \mathbf{P}_i et le vecteur \mathbf{q} telle que

$$\dot{\mathbf{P}}_i = [\mathbf{J}_g(\mathbf{q})]_i \dot{\mathbf{q}} \quad (1)$$

\mathbf{J}_g est une matrice Jacobienne géométrique ($6n \times 3$), et chaque sous-matrice (3×3) $[\mathbf{J}_g(\mathbf{q})]_i$ peut être décomposée en blocs \mathbf{A}_{q_k} reliant la variation d'un ddl q_k avec celles de la position 3-D d'un point \mathbf{P}_i selon l'expression

$$\dot{\mathbf{P}}_i = [\mathbf{A}_\lambda \mid \mathbf{A}_\phi \mid \mathbf{A}_\alpha]_i \begin{bmatrix} \dot{\lambda} \\ \dot{\phi} \\ \dot{\alpha} \end{bmatrix}^T \quad (2)$$

où $\mathbf{A}_{q_k} = \partial \mathbf{P}_i / \partial q_k$, et q_k est alternativement λ, ϕ ou α . A l'aide du modèle de projection perspective, la relation entre $\dot{\mathbf{P}}_i$ et le déplacement du point image correspondant $\dot{\hat{\mathbf{p}}}_i$ peut se mettre sous la forme $\dot{\hat{\mathbf{p}}}_i = \mathbf{J}_I(\mathbf{p}_i) \dot{\mathbf{P}}_i$, ce qui finalement s'écrit synthétiquement par

$$\dot{\hat{\mathbf{p}}}_i = \mathbf{J}_I(\mathbf{p}_i) [\mathbf{J}_g(\mathbf{q})]_i \dot{\mathbf{q}} \quad (3)$$

où $\mathbf{J}_I(\mathbf{p}_i) [\mathbf{J}_g(\mathbf{q})]_i = \mathbf{J}_i(\mathbf{p}_i, \mathbf{q})$. En empilant toutes les matrices \mathbf{J}_i , ($i = 1, \dots, 2n$), on obtient la matrice globale \mathbf{J} de dimensions ($4n \times 3$), dans la relation $\dot{\hat{\mathbf{p}}} = \mathbf{J} \dot{\mathbf{q}}$.

La variation des variables articulaires peut alors être calculée en suivant un schéma d'optimisation. A partir d'une solution initiale $\mathbf{q}^{(0)}$, et en suivant une approche de type Gauss-Newton, la correction à apporter à $\mathbf{q}^{(s-1)}$ pour résoudre itérativement (ici à l'étape (*step*) s) le problème posé (et à partir d'une seule acquisition d'image) consiste à minimiser $\chi_c^2 = \frac{1}{2} \sum_{i=1}^{2n} w_{ii}^2 \|\mathbf{p}_i - \hat{\mathbf{p}}_i\|^2$, et amène à

$$\mathbf{q}^{(s)} - \mathbf{q}^{(s-1)} = c \mathbf{J}_W^\dagger \mathbf{e}^{(s)} = c (\hat{\mathbf{J}}^T \mathbf{W}^T \mathbf{W} \hat{\mathbf{J}})^{-1} \mathbf{W}^T \mathbf{W} \mathbf{e}^{(s)} \quad (4)$$

où $\mathbf{e}^{(s)} = \mathbf{p} - \hat{\mathbf{p}}^{(s)}$ et \mathbf{J}_W^\dagger est la pseudo-inverse généralisée d'une estimation de la matrice $\hat{\mathbf{J}}$. c est un scalaire et \mathbf{W} est une matrice de pondération. Dans [20], \mathbf{W} est une matrice diagonale $\mathbf{W} = \text{diag}\{w_{ii}\}$ composée de 1 excepté pour les composantes relatives aux coordonnées des points supérieurs et inférieurs de l'extrémité de l'instrument ($i \in \{n, 2n\}$) qui sont égales à $\sqrt{10}$.

L'analyse effectuée sur plusieurs vidéos d'opérations chirurgicales réalisées à l'aide de la plateforme Anubis montre que les mouvements des instruments ont une faible amplitude entre deux images successives. Par conséquent la solution approchée pour le vecteur \mathbf{q}_f^* sur l'image f est utilisée directement comme solution initiale dans le même processus itératif exécuté sur l'image suivante, soit $\mathbf{q}_f^* = \mathbf{q}_{f+1}^{(0)}$. Dans l'approche de Reilink *et al.*, $\hat{\mathbf{p}}$ est calculé en supposant que la position et l'orientation du canal de sortie sont parfaitement connues. En appliquant cette technique à notre plateforme, et en s'appuyant sur son modèle CAO pour connaître la valeur des paramètres de la transformation géométrique entre la caméra et le canal de sortie (paramètres mécaniques), il s'avère que l'image calculée et correspondant à la meilleure configuration (selon la fonction de coût χ_c^2) ne coïncide pas avec l'image endoscopique originale (voir Fig. 4(a)). Il est à noter que nous avons aussi exécuté l'algorithme de minimisation décrit ci-dessus avec d'autres valeurs initiales que $\mathbf{q}_f^* = \mathbf{q}_{f+1}^{(0)}$ et que ceci ne semble pas être dû à la présence de minima locaux.

Avec une séquence de 116 images endoscopiques, les erreurs RMS relevées sont égales à 2.75 ± 2.36 , 3.61 ± 3.1 et 5.2 ± 3.06 mm respectivement pour les coordonnées x (horizontal), y (vertical) et z de l'extrémité de l'instrument.

Ces résultats suggèrent de revoir la pertinence de l'hypothèse portant sur la connaissance d'un modèle fiable

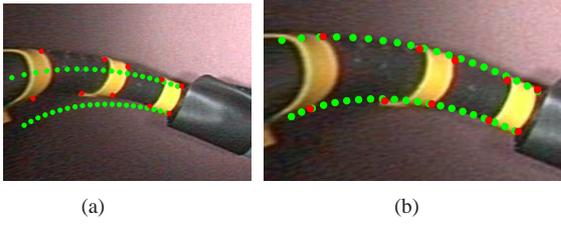


FIGURE 4 – Reprojection des bords de l’instrument (spots verts) à l’issue du processus d’optimisation. Les indices visuels sur l’image de référence sont les coins apparents (spots rouges). A gauche, avec les paramètres mécaniques constants issus du modèle CAO de la plateforme. A droite, avec une tolérance sur ces même paramètres mécaniques.

de l’ensemble (à noter que les indices visuels ont été sélectionnés manuellement ici) et particulièrement sur la constance des paramètres mécaniques. Bien que cette méthode soit intéressante pour estimer la pose d’un instrument flexible elle ne fournit pas une solution suffisamment précise en l’état. La prochaine section va revisiter cet algorithme en y adjoignant une variabilité maîtrisée sur les paramètres mécaniques.

3.2 Pose de l’instrument avec paramètres mécaniques variables

En incorporant les trois paramètres mécaniques correspondant à la position (x_{ch}, y_{ch}) et à l’orientation ψ du canal de sortie dans le processus d’optimisation présenté précédemment, un nouveau vecteur d’état $\mathbf{q}_e = [\mathbf{q}_m^T, \mathbf{q}^T]^T$ est défini où $\mathbf{q}_m = [x_{ch}, y_{ch}, \psi]^T$. La coordonnée z_{ch} du canal de sortie est maintenue constante et égale à sa valeur nominale, puisque la considérer variable reviendrait à créer une dépendance avec le ddl λ qui représente déjà un mouvement de translation selon cette direction.

En considérant que les variations restent confinées autour des valeurs nominales $\mathbf{q}_m^* = [x_{ch}^*, y_{ch}^*, \psi^*]^T$, des fonctions de pénalisation sont ajoutées dans la fonction de coût initiale afin de contenir ces valeurs. La nouvelle fonction de coût χ_v^2 prend alors la forme $\chi_v^2 = \chi_c^2 + \rho(\mathbf{q}_m^* - \hat{\mathbf{q}}_m)$ avec

$$\rho(\mathbf{q}_m^* - \hat{\mathbf{q}}_m) = \rho_d(x_d^* - \hat{x}_d) + \rho_d(y_d^* - \hat{y}_d) + \rho_\psi(\psi^* - \hat{\psi}). \quad (5)$$

Les fonctions de pénalisation ρ doivent attribuer un coût faible pour une valeur estimée d’un paramètre mécanique proche de sa valeur nominale et un coût de plus en plus élevé au fur et à mesure qu’elle s’en éloigne. Elles doivent de plus être dérivables, $\rho(x) \in \mathcal{C}^1$, dans tout le domaine ($x \in \mathbb{R}$), afin de pouvoir contribuer efficacement au processus d’optimisation. Un exemple d’une telle fonction est :

$$\rho(u) = \frac{k_w}{3a^3}|u|^3, \quad a \in \mathbb{R}^+, \quad k_w \in \mathbb{R}^+$$

avec la fonction de pondération correspondante :

$$\zeta(u) = \sqrt{\frac{k_w}{a^3}}|u|, \quad a \in \mathbb{R}^+, \quad k_w \in \mathbb{R}^+,$$

où k_w définit une pente abrupte en dehors de la zone morte, elle-même caractérisée par le paramètre a . Le nouveau vecteur d’erreur est alors le suivant :

$$\mathbf{e} = [(\mathbf{q}_m^* - \hat{\mathbf{q}}_m)^T \quad (\mathbf{p}_u - \hat{\mathbf{p}}_u)^T \quad (\mathbf{p}_l - \hat{\mathbf{p}}_l)^T]^T,$$

et l’équation (2) devient maintenant :

$$\dot{\mathbf{P}}_i = [\mathbf{A}_{x_{ch}} \mid \mathbf{A}_{y_{ch}} \mid \mathbf{A}_\psi \mid \mathbf{A}_\lambda \mid \mathbf{A}_\phi \mid \mathbf{A}_\alpha]_i \dot{\mathbf{q}}_e. \quad (6)$$

Le nouveau jacobien est de dimensions $((3 + 4n) \times 6)$, avec n points supérieurs et n points inférieurs délimitant les bords de chaque région marquée sur l’instrument. Finalement, la matrice de pondération, \mathbf{W}_v est

$$\mathbf{W}_v = \left[\begin{array}{ccc|c} \zeta(e_1) & 0 & 0 & \mathbf{O}_{3 \times 3} \\ 0 & \zeta(e_2) & 0 & \\ 0 & 0 & \zeta(e_3) & \\ \hline & \mathbf{O}_{24 \times 3} & & \mathbf{W} \end{array} \right].$$

Le schéma de résolution est le même que dans la section 3.1, basé sur l’équation (4), avec les nouvelles matrices et le nouveau vecteur d’erreur. Les résultats obtenus sur le dispositif expérimental montrent une bien meilleure précision dans la reprojection des points 3-D estimés dans la plupart des configurations. Un exemple d’une telle configuration est affiché sur la Fig. 4(b), à la convergence du processus d’optimisation avec des paramètres mécaniques variables et la nouvelle fonction de coût χ_v^2 . En comparant avec les données des capteurs électromagnétiques, l’erreur RMS est maintenant de 2.73 ± 1.04 , 1.21 ± 0.73 and 3.86 ± 2.21 mm respectivement pour les coordonnées x , y et z .

Une analyse approfondie de ces résultats montre qu’à l’issue du processus d’optimisation, les valeurs des composantes du vecteur \mathbf{q}_m varient en fonction des autres paramètres. En fait, au travers des 116 images analysées, la moyenne μ et l’écart-type σ des paramètres mécaniques sont égaux à :

- $\mu_x = E[x_{ch}] = -13.37$ mm et $\sigma_x = 1.091$ mm,
- $\mu_y = E[y_{ch}] = 3.863$ mm et $\sigma_y = 0.64$ mm,
- $\mu_\psi = E[\psi] = -7.86^\circ$ et $\sigma_\psi = 1.39^\circ$.

Avec des fluctuations maximales sur les valeurs moyennes de 2.1 mm, 1.65 mm et 3.86° respectivement pour x_{ch} , y_{ch} et ψ , on peut en conclure qu’il n’y a pas à proprement parler de valeurs fixes pour ces paramètres, car si on considère les valeurs moyennes comme les valeurs nominales et constantes et qu’on applique le schéma de la section 3.1, l’erreur RMS augmente. Comme la plupart des plateformes endoscopiques robotisées, Anubis est conçu avec un jeu important (au regard des dimensions des instruments) entre l’instrument flexible et les bords du canal, ce qui autorise un léger glissement et une petite rotation de l’instrument dans le canal. On en déduit que l’apport d’une tolérance sur les paramètres mécaniques correspond aussi bien à compenser des incertitudes sur l’ensemble du modèle qu’à prendre en compte les petits déplacements inévitables à l’intérieur du canal.

Ces valeurs statistiques cachent des configurations où les erreurs sur le positionnement 3-D sont plus importantes. Qu'il s'agisse des imprécisions sur la détection des indices visuels (qui ne sont pas exactement attachés à l'instrument ici), où bien sur le profil (vraiment circulaire ?) de la partie flexible, tout ceci contribue à engendrer des erreurs résiduelles, encore importantes pour une utilisation fiable en chirurgie. Pour améliorer cela, nous sommes en train d'étudier comment des indices visuels plus saillants pourraient réduire un peu plus les imprécisions sur le positionnement de l'instrument flexible. Mais parallèlement, nous avons aussi abordé ce problème de positionnement en explorant une technique d'apprentissage.

4 Méthode basée sur l'apprentissage

La méthode que nous proposons ici repose sur l'apprentissage d'une fonction qui relie les indices visuels (centres de gravité des marqueurs imagés) avec la position de l'instrument. Sa mise en oeuvre est composée de deux étapes : la segmentation puis l'estimation de la position 3-D.

4.1 Segmentation de l'instrument

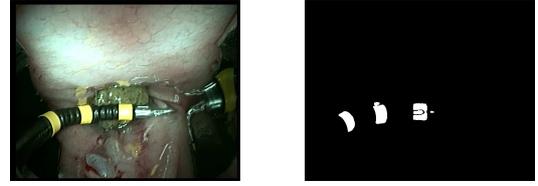
Détection des indices visuels. Qu'il s'agisse de la flexion de l'instrument ou des variabilités de l'arrière-plan en condition *in vivo*, une segmentation basée sur la forme ou sur la détection d'indices visuels qu'offre l'apparence de n'importe quel instrument est un problème difficile. L'ajout de marqueurs artificiels colorés permet la mise en oeuvre d'une approche spectrale pour la segmentation. Une classification par la couleur est alors réalisée grâce à un classifieur adaptatif, comme celui qui offre l'algorithme Adaboost [21, 22]. La segmentation peut alors être vue comme un problème de classification des couleurs, où des classifieurs faibles h_k sont définis et combinés pour former un classifieur fort, H_N , de la manière suivante :

$$H_N(x) = \text{sign} \left(\frac{\sum_{k=1}^N \alpha_k h_k(x)}{\sum_{k=1}^N \alpha_k} \right). \quad (7)$$

Dans cette relation, x est un échantillon de données à classer et α_k est la pondération attribuée au classifieur h_k estimé durant la phase d'apprentissage. Si la combinaison qui en résulte, H_N , est positive, alors l'échantillon x est étiqueté avec un statut d'indice visuel.

Une analyse en composantes principales est effectuée ici sur les données à entraîner, pour mettre en oeuvre la classification des échantillons dans un espace propre. Cet espace tridimensionnel des couleurs est alors subdivisé en sous-espaces délimités par des plans perpendiculaires à l'axe principal. Ces plans servent à définir le domaine d'action de chaque classifieur faible h_k .

A partir d'un petit échantillon d'images *in vivo*, l'entraînement est réalisé en sélectionnant manuellement des éléments de l'arrière-plan (sur différentes parties internes d'un porc) et sur les indices visuels à classer. Un exemple de l'application d'un tel classifieur sur une image test est affiché sur la Fig. 5. Comme mentionné précédemment, l'am-



(a) Image endoscopique *in vivo*.

(b) Segmentation.

FIGURE 5 – Exemple d'une image *in vivo* de l'instrument flexible par la caméra endoscopique (a) et segmentation des marqueurs obtenue par le classifieur Adaboost (b).

plitude des déplacements des instruments est limitée entre deux images successives. Dès l'obtention d'une segmentation de l'instrument sur la première image, le classifieur peut agir le long de la séquence sur une région très limitée. Au-delà de la réduction du temps de traitement, ceci permet d'éviter la détection de faux positifs, comme par exemple les réflexions des marqueurs sur les organes visuels exhibant des surfaces brillantes.

Étiquetage des indices visuels. Etant donné la structure de robot continu que présente un instrument flexible, il est raisonnable de considérer que l'orientation de l'instrument en sortie du canal ne change pas significativement (aux incertitudes près évoquées précédemment et au bruit près sur les images). Ceci signifie que la zone où la projection de l'instrument rencontre l'image (de la caméra embarquée) est à peu près toujours la même, correspondant à une région quasi-statique située aux bords de celle-ci. Cette particularité que l'on retrouve pour la laparoscopie peut servir à rechercher le premier indice visuel, image du marqueur placé à l'interface entre la partie rigide et la partie flexible. Ce premier indice visuel est utilisé par la suite pour délimiter les régions correspondant aux autres indices visuels. Ainsi, une opération de dilatation binaire $D_d(\cdot)$ d'ordre d peut alors être réalisée à l'aide d'un élément structurant de dimensions en lien avec la taille de la région englobant le premier indice visuel détecté. Nous utilisons pour cela les *bords étendus*, similaire au gradient par dilatation, qui est défini par le résultat de la soustraction de l'image binaire dilatée à l'ordre d par celle à l'ordre $d - 1$, $B_d(I) = D_d(I) - D_{d-1}(I)$, où I est l'image binaire. Les limites de la région associée au premier indice visuel sont déterminées de cette manière et l'intersection entre cette frontière et l'image segmentée est calculée, permettant d'étiqueter les pixels de l'image s'y trouvant et de les différencier de ceux appartenant aux autres indices visuels.

4.2 Estimation de la position 3-D

La fonction qui relie la position 3-D de l'extrémité de l'instrument avec les coordonnées des (centre des) indices visuels des images doit être identifiée. Pour ce problème de régression nous utilisons un réseau de fonctions à base radiale (RBF), évitant ainsi les oscillations que

créerait une interpolation polynomiale. De telles fonctions, construites à partir de noyaux gaussiens, reflètent bien les variations locales et peuvent être considérées comme formant un réseau connexionniste à deux couches [23, 24]; La première couche est formée par les noyaux gaussiens qui sont calculés à l'aide d'un apprentissage non supervisé. La seconde consiste à obtenir la meilleure combinaison linéaire de ces noyaux pour représenter la fonction recherchée, à partir d'un apprentissage supervisé.

Soient $\mathbf{X}^f = (x_1^f, \dots, x_{2n}^f)$ les coordonnées images des n centres de gravité des régions formant des indices visuels et soient $\mathbf{Y}^f = (y_1^f, y_2^f, y_3^f)$ les coordonnées de la position 3-D de l'extrémité de l'instrument flexible correspondant à la même image (image f). Chaque coordonnée de position est reconstruite, avec le réseau RBF, comme une somme pondérée de K noyaux judicieusement choisis :

$$y_i^f = \sum_{j=1}^K \mathcal{W}_i^j \varphi^j(\mathbf{X}^f) + \mathcal{W}_i^{K+1} \quad (8)$$

où \mathcal{W}_i^j sont des coefficients appris. Dans le cas de noyaux gaussiens, φ^j prend la forme suivante [25] :

$$\varphi^j(\mathbf{X}^f) = \mathcal{Y}^j e^{-\frac{1}{2}(\mathbf{X}^f - \mathbf{C}_j)' \Sigma_j^{-1} (\mathbf{X}^f - \mathbf{C}_j)} \quad (9)$$

où \mathbf{C}_j est le centre du noyau j , Σ_j sa variance et \mathcal{Y}^j son amplitude. Un apprentissage sur N données permet alors de calculer les valeurs des paramètres du réseau.

La recherche de la meilleure combinaison des centres des noyaux nécessite de choisir aléatoirement K échantillons de données. Ces échantillons servent à initialiser des grappes (*clusters*) qui évolueront à travers un algorithme de type k -moyennes agissant sur l'ensemble des données. Quand les grappes sont constituées (plus d'évolutions des K centres), les variances et les amplitudes caractérisant chaque grappe sont calculées. Les pondérations sont finalement obtenues par la minimisation de l'erreur résiduelle entre les données de sortie de l'ensemble d'apprentissage et ce qui a été évalué à travers le réseau RBF, soit :

$$\min_{\mathcal{W}_i} \left(\sum_{f=1}^N (y_i^f - \mathcal{W}_i^{K+1} - \sum_j \mathcal{W}_i^j \varphi^j(\mathbf{X}^f)) \right).$$

Les résultats obtenus, à partir de la séquence des 116 images utilisées aussi par les méthodes basées modèle, sont représentés sur la Fig. 6. On voit clairement que la précision obtenue par cette nouvelle méthode surpasse les méthodes basées modèle la plupart du temps.

5 Conclusion

Dans cet article l'estimation de la position 3-D d'un instrument de chirurgie flexible robotisé à partir d'une vision endoscopique a été étudiée. L'incertitude de plusieurs paramètres du modèle cinématique de l'ensemble du robot et la difficulté de segmenter des images d'un environnement *in*

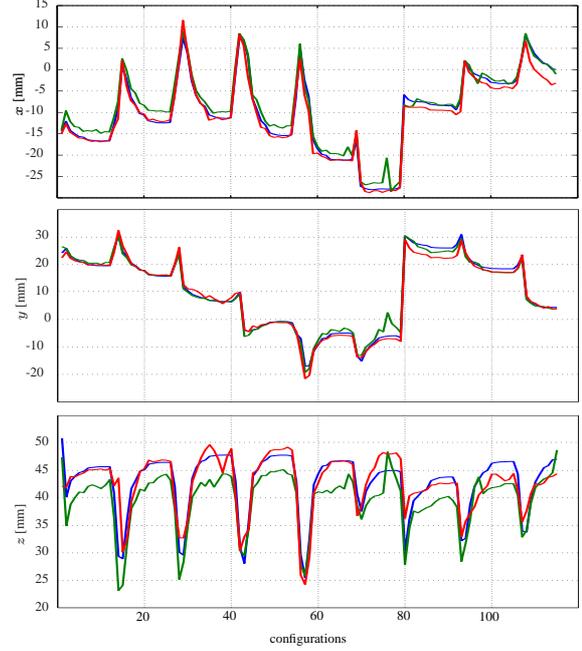


FIGURE 6 – Comparaison entre la position 3-D de référence fournie par le capteur électromagnétique (en bleu), l'estimation de cette position par la méthode d'apprentissage (en rouge) et l'estimation par la méthode basée sur le modèle de l'instrument avec paramètres mécaniques variables (en vert).

vivo amène à une imprécision forte sur la position de l'extrémité et sur la configuration réelle de l'instrument flexible tout le long de la chaîne cinématique. Pour surmonter cela, nous avons proposé deux méthodes pour améliorer la précision de cette estimation. La première méthode a consisté à revisiter la technique de Reilink *et al.* en autorisant une variation modérée de certains paramètres du modèle. Les résultats obtenus sont encourageants, et montrent, en utilisant une segmentation manuelle des points d'intérêt, une amélioration notable de la précision. La seconde est une nouvelle technique qui est basée sur un apprentissage supervisé, et qui résout le problème délicat de la segmentation automatique des images endoscopiques.

Le choix entre les différentes approches présentées reste à discuter. Des précisions comparables sont obtenues entre la méthode basée sur l'apprentissage et la méthode basée modèle avec paramètres mécaniques variables. D'un côté, la simplicité des traitements des images nécessaires à la méthode par apprentissage est un avantage par rapport à la difficulté d'extraire automatiquement et avec précision les indices visuels nécessaires aux méthodes basées modèle. De l'autre, la méthode par apprentissage requiert de suivre un schéma de mise en oeuvre pratique plus contraignant que les méthodes basées modèle ; En particulier le flux de traitement est plus complexe et un capteur supplémentaire est nécessaire pour collecter des données vérité terrain.

Références

- [1] R. Zorrón, M. Filgueiras, L. C. Maggioni, L. Pombo, G. Lopes Carvalho, and A. Lacerda Oliveira, "NOTES. transvaginal cholecystectomy : report of the first case," *Surg Innov*, vol. 14, no. 4, pp. 279–283, Dec. 2007.
- [2] A. De Donno, L. Zorn, P. Zanne, F. Nageotte, and M. de Mathelin, "Introducing STRAS : a new flexible robotic system for minimally invasive surgery," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 1213–1220.
- [3] B. Jones and I. Walker, "Kinematics for multisection continuum robots," *IEEE Transactions on Robotics*, vol. 22, no. 1, pp. 43–55, 2006.
- [4] R. W. III, J. Romano, and N. Cowan, "Mechanics of precurved-tube continuum robots," *IEEE Transactions on Robotics*, vol. 25, no. 1, pp. 67–78, 2009.
- [5] V. Agrawal, W. Peine, B. Yao, and S. Choi, "Control of cable actuated devices using smooth backlash inverse," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 1074–1079.
- [6] T. Reichl, J. Gardiazabal, and N. Navab, "Electromagnetic servoing - a new tracking paradigm," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 1526–1535, 2013.
- [7] B. Bardou, P. Zanne, F. Nageotte, and M. de Mathelin, "Control of a multiple sections flexible endoscopic system," in *2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2010, pp. 2345–2350.
- [8] M. Hannan and I. Walker, "Vision based shape estimation for continuum robots," in *IEEE International Conference on Robotics and Automation*, vol. 3, 2003, pp. 3449–3454.
- [9] V. Chitrakaran, A. Behal, D. Dawson, and I. Walker, "Setpoint regulation of continuum robots using a fixed camera," in *Proc. of the American Control Conference*, vol. 2, 2004, pp. 1504–1509.
- [10] D. B. Camarillo, K. E. Loewke, C. R. Carlson, and J. K. Salisbury, "Vision based 3-D shape sensing of flexible manipulators," in *Robotics and Automation. ICRA, IEEE International Conference on*, 2008, pp. 2940–2947.
- [11] B. Weber, P. Zeller, and K. Kuhnlenz, "Multi-camera based real-time configuration estimation of continuum robots," in *IEEE/RSJ Intelligent Robots and Systems*, 2012, pp. 3350–3355.
- [12] F. Nageotte, P. Zanne, C. Doignon, and M. D. Mathelin, "Stitching planning in laparoscopic surgery : towards robot-assisted suturing," *International Journal of Robotics Research*, vol. 28, no. 10, pp. 1373–1386, October 2009.
- [13] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1050–1058, 2013.
- [14] R. Reilink, S. Stramigioli, and S. Misra, "3D position estimation of flexible instruments : marker-less and marker-based methods," *Int. Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 3, pp. 407–417, 2013.
- [15] T. Hu, P. Allen, T. Nadkarni, N. Hogle, and D. Fowler, "Insertable stereoscopic 3D surgical imaging device with pan and tilt," in *2nd IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechanics, BioRob.*, 2008, pp. 311–316.
- [16] A. Reiter, R. E. Goldman, A. Bajo, K. Iliopoulos, N. Simaan, and P. Allen, "A learning algorithm for visual pose estimation of continuum robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 2390–2396.
- [17] J. C. K. Chou and M. Kamel, "Quaternions approach to solve the kinematic equation of rotation, $ax=xb$, of a sensor-mounted robotic manipulator," in *IEEE International Conference on Robotics and Automation*, 1988, pp. 656–662.
- [18] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotic Research*, vol. 18, no. 14, pp. 286–298, 1999.
- [19] E. Marchand and F. Chaumette, "Virtual visual servoing : a framework for real-time augmented reality," in *Eurographics : Computer Graphics Forum*, vol. 21, no. 3, 2002, pp. 289–297.
- [20] R. Reilink, S. Stramigioli, and S. Misra, "Pose reconstruction of flexible instruments from endoscopic images using markers," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2012, pp. 2938–2943.
- [21] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119–139, August 1997.
- [22] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin : A new explanation for the effectiveness of voting methods," *The Annals of Statistics*, vol. 26, no. 5, pp. 1651 – 2080, 1998.
- [23] D. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.
- [24] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, no. 2, pp. 281–294, 1989.
- [25] A. G. Bors, "Introduction of the radial basis function (rbf) networks," in *Online Symposium for Electronics Engineers*, 2001.