

# Suivi 3D multi-objets basé modèle par vidéo pour la surveillance de salles de traitement

M.A. Portela Sotelo<sup>1,2</sup>

É. Desserée<sup>1</sup>

J-M. Moreau<sup>1</sup>

<sup>1</sup> Université Lyon1, LIRIS, UMR5205, F-69622, France

<sup>2</sup> DOSIsoft S.A., Cachan, France.

Université Lyon 1, Campus de la DOUA, 69622 Villeurbanne Cedex, France

e-mail : miguel.portela-sotelo@liris.cnrs.fr

## Résumé

*Nous présentons une méthode pour le suivi de mouvements d'un patient et de l'équipement d'une salle de traitement en radiothérapie, qui exploite des données contenues dans le plan de traitement, enrichies d'autres informations visuelles, géométriques et « sémantiques ». Un modèle générique, que nous avons développé, permet de créer un environnement virtuel à partir de ces informations. En utilisant également des images acquises à partir de plusieurs caméras, la pose des éléments est calculé par recalage 3D, avec une fonction de dissimilarité que nous avons défini. La méthode est fondée sur les caractéristiques des éléments, et sur un formalisme flou. Les premiers tests en l'absence de traitement démontrent la faisabilité de la méthode, qui sera exploitée dans un module de sécurité par contrôle externe.*

## Mots Clef

Réalité augmentée et virtuelle, vision par ordinateur, analyse vidéo, suivi de mouvements, segmentation, évaluation de caractéristiques, interactivité, logique floue.

## Abstract

*We present a method to track a patient and the equipment in a radiotherapy treatment room, by exploiting the information in the treatment plan, enriched with other elements such as visual, geometric and « semantic » information. A generic model, that we developed, is used to create a virtual environment from these information items. Also using images resulting from several cameras, the pose of the elements is calculated by 3D registration, with a dissimilarity function we defined. The method is based on the features of the elements, and on a fuzzy formalism. The firsts tests in the absence of treatment demonstrated the feasibility of the method, which will be exploited in an security module by external control.*

## Keywords

Augmented, and virtual reality, Computer vision, Video analysis, Tracking, Feature evaluation and selection, Interactive systems, Fuzzy logic.

## 1 Contexte

**La réalité augmentée dans les systèmes médicaux** La modélisation 3D [+ temps] du patient à partir des images TDM (TomoDensitoMétrie) ou IRM (Imagerie par Résonance Magnétique) permet aujourd'hui de mieux planifier l'acte chirurgical, grâce à une meilleure connaissance du patient. Une version numérique précise et fiable de l'anatomie du patient offre la possibilité de s'entraîner aux gestes opératoires par une simulation personnalisée de l'opération, et donc de réduire sensiblement les erreurs médicales. Une autre utilisation d'un tel patient virtuel est le suivi temporel, permettant de visualiser plus facilement l'efficacité d'une thérapie. La réalité augmentée permet un guidage plus efficace du geste opératoire, en superposant au patient les informations pré-opératoires. Cette possibilité est rarement exploitée par les simulateurs actuels [1, 2, 3].

**Le contrôle externe en radiothérapie** Il est actuellement possible de trouver dans les salles de traitement standard des systèmes d'aide au positionnement précis du patient : par exemple AlignRT, Polaris, ou des accessoires d'équipements d'irradiation comme CyberKnife, Novalis. Des modules de contrôle externe pour la radiothérapie, qui visent l'automatisation du positionnement du patient, ont été également proposées [4, 5, 6]. Bien que ces dispositifs soient focalisés sur le patient et ne prennent pas en compte son environnement, ils attestent de l'importance des systèmes de contrôle externe en radiothérapie.

## 2 Objectifs

Lors de la création du plan de traitement radiothérapeutique pour un patient, un ensemble de configurations est prédéfini. Cet ensemble contient notamment une configuration initiale, et des configurations pour les différentes orientations dans lesquelles la tumeur sera irradiée. Dans l'environnement virtuel de la scène, il est possible de générer l'équivalent de la configuration initial. Les paramètres de cette configuration sont supposés invariants, sauf lorsqu'il est besoin de se placer dans une autre configuration prédéfinie (*c'est-à-dire* changer d'orientation de l'irradia-

tion). L'objectif principal de la surveillance de la salle de traitement est alors de vérifier que cette hypothèse d'invariance est toujours satisfaite et conforme au plan de traitement, ainsi qu'aux normes de sécurité établies pour ce type de traitement - le personnel médical devant être alerté chaque fois qu'on s'approche d'une situation critique ou risquée.

Notre travail met l'accent sur les points suivants :

1. la détection de déplacements non programmés (par rapport aux instructions du plan de traitement et un seuil de « distance » donné), soit du patient, soit des objets ou équipements de la salle,
2. la détection de personnes ou d'objets dont la présence est non planifiée au cours de la session,
3. le respect des configurations prédéfinies, ainsi que le respect des normes de sécurité lors de transitions entre les différentes configurations.

## 3 Une approche générique

### 3.1 Le modèle numérique

De nombreux travaux se sont focalisés sur le développement de méthodes pour la simulation numérique des comportements physiques, dans le but de les intégrer dans des simulateurs médicaux [7, 8]. Les modèles numériques que ces méthodes utilisent sont souvent le résultat de la segmentation, manuelle ou assistée, des données issues de l'imagerie, et n'ont pas un caractère adaptatif ni générique. Nous utilisons le modèle générique que nous avons développé [9] pour répondre, principalement, à deux besoins :

- l'adaptabilité pour tout nouvel élément à ajouter à la scène, avec la capacité de fournir une instance numérique aussi précise que possible de cet élément.
- la généralité pour tout type d'information nécessaire aux différentes méthodes exploitées par le module.

La complexité de la tâche impose de modéliser la totalité des éléments de la salle de traitement (individus et objets). Plus la quantité d'informations est élevée, plus on est capable de lever les ambiguïtés liées au système d'acquisition (occultation totale ou partielle, éclairage, etc.).

Notre modèle s'appuie sur les principes de la Géométrie de Construction de Solides (CSG). Une telle structure, arborescente, est bien adaptée à une représentation des objets décomposables et permet facilement la sauvegarde des informations en utilisant un langage à balises, comme le XML. Les éléments sont donc associés par des opérateurs pouvant décrire une interaction physique, géométrique et/ou sémantique. Par exemple, une porte est attachée à un mur (interaction physique), elle s'ouvre uniquement vers l'intérieur de la salle, bute au contact du mur attenant (interaction géométrique), et elle ne s'ouvre si un autre élément est assez proche pour la pousser ou la tirer (interaction sémantique). Les opérateurs permettent également de regrouper différentes représentations d'un même élément.

Les *paramètres* d'un élément dans le modèle définissent ses propriétés, comme la couleur, les dimensions, la forme, les paramètres mécaniques, les configurations « cibles », les statistiques, les éléments associés, etc.. Une partie de ces paramètres a comme source des fichiers d'autres formats spécifiques : DICOM<sup>1</sup>, VRML<sup>2</sup>, par exemple.

Le point fort de ce modèle repose sur la notion de *descripteur*. Un descripteur est une caractéristique ou un objet associé à l'élément qui permet de le reconnaître visuellement (mais pas forcément de manière unique). Cela peut être sa couleur uniforme, ou bien un motif déjà présent, des gommettes de couleur, ou bien des balles à réflexion d'infrarouge sur la surface de l'élément. Un descripteur est donc une représentation numérique dans l'espace 3D qui, par projection dans le plan image des caméras en fonction des paramètres intrinsèques et extrinsèques, correspond à des primitives 2D (contours, points caractéristiques, motifs, texture, etc.). La « pose » 3D d'un descripteur par rapport à l'élément doit être connue et fixe, car elle représente un référentiel de l'élément : si la pose du descripteur a pu être calculée par rapport au repère global de la scène, celle de l'élément par rapport au repère global est également disponible car la transformation mathématique entre les deux repères est une donnée connue.

Finalement, à chaque objet lui est associé un maillage triangulaire qui représente sa surface externe. Cet maillage est découpé en fonction des descripteurs et des parties indépendantes de l'élément (degrés de liberté).

### 3.2 Suivi d'individus et d'objets

Nous avons décidé de nous concentrer sur les techniques basées modèle, afin d'exploiter la notion de patient virtuel. Ces techniques sont connues pour être plus précises et robustes car le modèle fournit un grand nombre d'informations à tout instant et aux différentes étapes de la méthode. Bien qu'elles soient rarement envisageables pour des systèmes à temps interactif, à cause de leur complexité en volume de calculs, elles sont particulièrement adaptées à nos besoins, principalement à cause de l'exploitation des contraintes liées au contexte médical, offrant un environnement théoriquement sous contrôle, et qui permet de réduire de manière significative les temps d'exécution.

## 4 La solution proposée

L'aspect multi-primitives, apporté par la notion de descripteur 3D, rend plus précis et robustes les résultats de la détection et du suivi de mouvements sur les images vidéo [10]. Nous exploitons également les contraintes mécaniques des éléments, et en particulier les degrés de liberté et les espaces de collision [11], tels que définis dans la scène. Enfin, une stratégie fondée sur la logique floue [12] permet de fusionner les informations fournies par tous les descripteurs pour définir les positions dans l'espace 3D de tous les éléments modélisés de la salle.

1. Digital Imaging and COmmunications in Medicine  
2. Virtual Reality Modeling Language

L'idée générale de la méthode est de permettre à deux environnements d'interagir : l'Environnement Réel, ER, (*c'est-à-dire*, la scène filmée par les caméras) et l'Environnement de Réalité Virtuelle, ERV, (construit en utilisant le modèle générique et contenant une description numérique de la scène : des humains et des objets).

Une représentation numérique du système d'acquisition, contenant les paramètres intrinsèques et extrinsèques des caméras, est également présent dans l'ERV. Le modèle numérique  $Vcam_i$  d'une caméra réelle  $Rcam_i$  permet de générer, à l'aide des paramètres, une image virtuelle  $Vimg_{(i,t)}$  équivalant à une image réelle  $Rimg_{(i,t)}$  acquise depuis la caméra  $i$  à tout temps  $t$ ,  $t$  étant lié à la fréquence d'acquisition du système de caméras synchronisées.  $Vimg_{(i,t)}$  est donc une simulation de l'acquisition d'une image  $Rcam_i$ . Après la segmentation des primitives 2D sur  $Rimg_{(i,t)}$  et  $Vimg_{(i,t)}$ , on veut corréler les informations contenues dans les deux images.

À un instant  $t$ , si la position d'un élément de la scène dans l'ER correspond exactement à celui de sa représentation numérique des ERV alors, pour toute caméra  $Rcam_i$ , la projection de l'élément sur le plan image dans l'ER est équivalente à la projection du modèle numérique de l'élément sur le plan image de  $Vcam_i$  dans l'ERV.

Ce principe d'identité est fréquent dans les méthodes de suivi vidéo basées modèle. Ces méthodes exploitent l'espace de solutions de positions possibles, afin de calculer la meilleure approximation. Plusieurs articles de recherche utilisent ces méthodes pour le suivi des personnes ou de parties de leur corps, en utilisant des modèles humanoïdes (volumiques ou surfaciques) [13, 14], mais aussi pour le suivi d'objets (avec des modèles rigides ou non), notamment appliqué à l'« asservissement visuel » [15, 16]. Ces recherches démontrent la faisabilité de la technique dans le domaine du suivi de mouvement, et leur grand avantage sur les autres méthodes publiées, lorsqu'on dispose d'une connaissance *a priori* de l'objet cible.

#### 4.1 Utilisation de descripteurs 3D

Nous ne sommes pas nécessairement intéressés par le suivi de l'élément tout entier, mais par celui des caractéristiques spécifiques qui sont moins difficiles à traiter sur les images vidéo. Les informations obtenues à partir des images sur quelques caractéristiques suffisent généralement à positionner l'élément grâce au modèle numérique.

Comme le montre le schéma de la Fig. 1, après une phase d'initialisation (création de l'environnement virtuel notamment), la méthode est constituée d'une boucle globale dont chaque itération correspond à une acquisition synchronisée du système de caméras.

Dans une itération, on retrouve également une deuxième boucle pour le calcul de la pose d'un élément, basée sur la génération de modèles numériques 2D des primitives associées aux descripteurs 3D de l'élément courant, pour chaque caméra et pour chaque pose évaluée. Une image  $VSimg_{(i,t,j,k)}$  générée contient la projection du descrip-

teur  $Desc_{(j,k)}$  de l'élément  $Elem_j$  sur le plan de l'image de  $Vcam_i$  au temps  $t$ . L'ensemble des pixels non-nuls dans  $VSimg_{(i,t,j,k)}$  constitue la primitive  $Vprim_{(i,t,j,k)}$ .

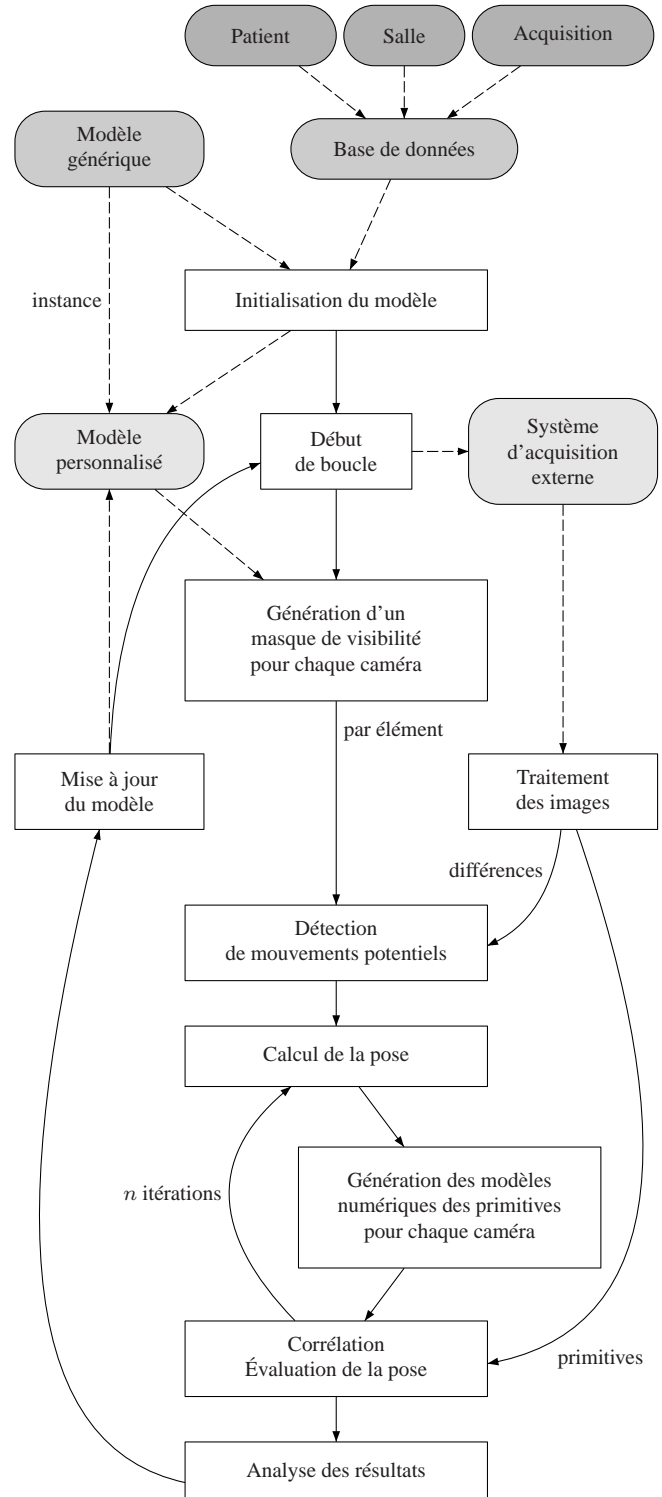


FIGURE 1 – Diagramme général de la méthode.

$RSimg_{(i,t,j,k)}$  est calculée en segmentant l'image  $Rimg_{(i,t)}$  selon le type de la primitive représentant  $Desc_{(j,k)}$  : par exemple, pour les contours, nous utilisons

un filtre de type Canny, pour les descripteurs colorimétriques, nous exploitons l'apprentissage d'histogramme HSV (Teinte Saturation Valeur) pour obtenir une carte de probabilité, etc.. Comme précédemment, l'ensemble des pixels non-nuls de l'image après segmentation constitue la primitive  $Rprim_{(i,t,j,k)}$ , dont les informations seront confrontées aux informations contenues dans  $Vprim_{(i,t,j,k)}$ .

Afin de réduire le volume de calculs, on n'évalue une nouvelle position que si un mouvement a été détecté. Pour ceci, on génère une image  $VSimg_{(i,t,j,k_v)}$ , pour chaque élément indépendamment, qui correspond au descripteur « plus petit volume englobant » et qui contient également la valeur de profondeur de chaque pixel (z-buffer). Si l'on considère tous les éléments, on est capable d'obtenir pour chaque élément une image binaire dont toutes les valeurs maximales appartiennent à la partie visible de l'élément (nommée « masque de visibilité »). Cette image représente la zone où l'élément était présent au temps  $t - 1$  (au niveau image). En calculant la différence entre  $Rimg_{(i,t)}$  et  $Rimg_{(i,t-1)}$ , et en utilisant l'image calculé comme un masque (au niveau de image) des pixels à traiter, il est possible de détecter un mouvement potentiel, que d'autres descripteurs pourront ensuite aider à quantifier.

## 4.2 La fonction de dissimilarité

Les méthodes de suivi basées sur un modèle peuvent également être classées en deux autres catégories : les méthodes stochastiques, et les méthodes déterministes. Les premières utilisent des techniques d'estimation comme les filtres de Kalman, filtres à particules, etc. [17]. La dynamique de la scène étant relativement faible d'une image à la suivante, en moyenne sur une séance de traitement type, de telles méthodes sont moins adaptées à nos besoins.

Les méthodes déterministes sont fondées sur la minimisation itérative d'une fonction de coût mesurant l'alignement du modèle sur des images réelles. Deux fonctions très connues, basées sur les contours des objets, sont la distance de chanfrein [18] et la distance de Hausdorff [19]. Du fait qu'elles sont très sensibles aux valeurs aberrantes, [11] a suggéré une fonction basée sur la mesure du non-recouvrement des surfaces. Nous proposons une fonction de dissimilarité qui fusionne la distance de chanfrein et la notion de non-recouvrement :

$$\rho(R, V) = \left(1 - \frac{|R \cap V|}{|R \cup V|}\right) \times \frac{1}{|CV|} \times \sum_{v_a \in CV} \min_{r_b \in CR} d(v_a, r_b) \quad (1)$$

où  $R = Rprim_{(i,t,j,k_d)}$ ,  $V = Vprim_{(i,t,j,k_d)}$ ,  $CV = CVprim_{(i,t,j,k_d)}$ ,  $CR = CRprim_{(i,t,j,k_d)}$ , et :

$$CVprim_{(i,t,j,k_d)} = Vprim_{(i,t,j,k_d)} \cap Vprim_{(i,t,j,k_c)} \quad (2)$$

où  $Vprim_{(i,t,j,k_c)}$  est la primitive contour de l'élément. Une formule analogue existe pour  $CRprim$ .

L'évaluation d'une pose fonctionne selon deux cas mutuellement exclusifs. Soient  $3Dpose_A$ ,  $3Dpose_B$  deux poses différentes contenues dans l'espace de solutions et différentes de la pose cible  $3Dpose_O$  :

1. lorsque la silhouette de l'élément projeté par rapport  $3Dpose_A$  a une intersection vide avec la silhouette cible  $3Dpose_O$  (voir la figure 2(a)), la première moitié de la formule ci-dessus  $\left(1 - \frac{|R \cap V|}{|R \cup V|}\right)$ , correspondant à la notion de non-recouvrement, prend la valeur 1, et la fonction se comporte comme une fonction de chanfrein :

$$\rho(R, V) = \frac{1}{|CV|} \times \sum_{v_a \in CV} \min_{r_b \in CR} d(v_a, r_b),$$

ce qui donne plus de poids à la translation pour la convergence ;

2. dans le cas contraire (Fig. 2(b)), la première moitié de la formule apporte des informations plus précises sur la rotation, car elle porte sur le nombre de pixels et non sur une moyenne de distances.

Utiliser la silhouette de la primitive permet de surmonter les problèmes liés aux occultations et aux variations d'éclairage. Utiliser les contours de la primitive, et non pas seulement les contours de la silhouette de la primitive, permet de garantir la robustesse lors d'occultations partielles et lorsque la silhouette est symétrique par rapport à un ou plusieurs degrés de liberté de l'élément.

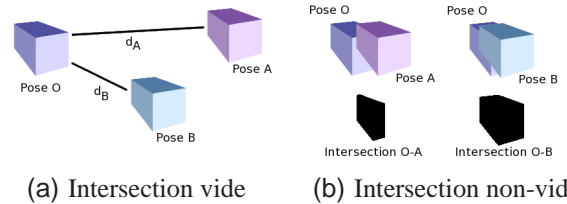


FIGURE 2 – Illustration de la fonction de dissimilarité.

## 4.3 Fusion des informations issues des primitives

Une pose 3D  $3Dpose$  est définie par six paramètres  $(x \ y \ z \ \Psi \ \theta \ \varphi)$ , qui correspondent à des translations et trois rotations dans l'espace 3D (angles d'Euler). Le point de départ pour la recherche est la position de l'objet au temps  $t - 1$ . L'espace de solutions est borné par les contraintes mécaniques et spatiales ajoutées au modèle et les degrés de liberté de l'élément.

Afin de pallier les problèmes liés aux occultations et au bruit généré par la manipulation d'images réelles, nous utilisons le principe de la logique floue : les ensembles flous permettent de fusionner des données hétérogènes et de prendre en compte l'imprécision générée par l'acquisition et le traitement des données.

Nous définissons deux ensembles flous (Fig. 3) correspondant à la visibilité (au niveau de l'image) et au déplacement dans l'espace 3D. Ces ensembles permettent d'associer une mesure de poids représentant un taux de confiance sur l'approximation calculée pour une pose cible.

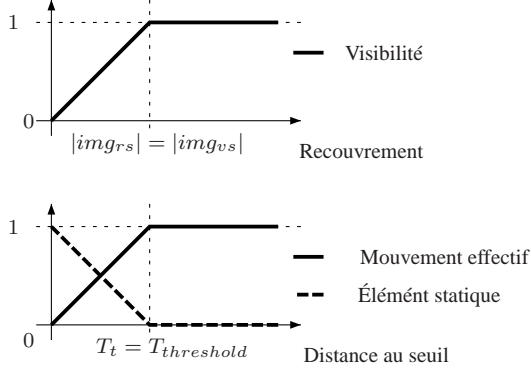


FIGURE 3 – Ensembles flous associés.

À chaque descripteur  $Desc_{(j,k)}$  d'un élément  $j$  est associé un poids  $Pdesc_{(j,k,t)}$ , qui résulte de la valeur de visibilité du descripteur, calculée par le recouvrement de la silhouette et de la partie visible de la silhouette, et la valeur donnée par la fonction de dissimilarité 1 :

$$Pdesc_{(j,k,t)} = \prod_i^n (1 - Val_{(i,t,j,k)}^v \times \rho(R, V)) \quad (3)$$

où  $Val_{(i,t,j,k)}^v$  est la valeur de visibilité de  $Vprim_{(i,t,j,k)}$ . La valeur de déplacement est calculée en utilisant la transformation  $T^{3D}$  entre  $3Dpose_{(j,t-1)}$  et  $3Dpose_{(j,t)}$ .  $T^{3D}$  est définie, par rapport à un vecteur seuil  $S = (x_s \ y_s \ z_s \ \Psi_s \ \theta_s \ \varphi_s)$  déterminé en fonction des contraintes de chaque élément, comme :

- en dessous du seuil : si tous ses composants sont strictement inférieurs à leurs homologues respectifs dans  $S$ ,
- au dessus du seuil : sinon.

Dans l'ensemble correspondant au déplacement nous avons défini deux états : *statique* et *mouvement effectif*, dont les valeurs sont données par une fonction d'évaluation entre  $T^{3D}$  et  $S$ .

Une pose  $3Dpose_{(j,t)}$  d'un élément  $Elem_j$  est évaluée par rapport au descripteur  $Desc_{(j,k)}$  dont le poids  $Pdesc_{(j,k,t)}$  est supérieur à un seuil prédéfini :

$$\sigma_{(t,j,k)} = \begin{cases} Pdesc_{(t,j,k)} & \text{if } Pdesc_{(t,j,k)} > S \\ 1 & \text{else} \end{cases} \quad (4)$$

Enfin, le poids  $P3Dpose_{(j,t)}$  associé à  $3Dpose_{(j,t)}$  est donné par :

$$P3Dpose_{(j,t)} = (2 - \max(Val_s, Val_m)) \times \prod_k^n \sigma_{(t,j,k)} \quad (5)$$

où  $Val_s$  et  $Val_m$  sont, respectivement, les valeurs des états statique et mouvement effectif.

#### 4.4 Calcul de la pose

L'exploration de l'espace de solutions est effectuée en utilisant l'approche fondée sur les simplexes proposée par Nelder et Mead [20]. Ce choix se justifie par la simplicité de la mise en œuvre et le fait que la dérivée de la fonction à minimiser n'est pas nécessaire. Cette approche pourrait également être parallélisée afin de réduire les temps de calcul, grâce au fait que l'espace est parcouru dans des directions différentes par la notion de simplexe, polytope de  $N + 1$  sommets dans un espace  $N$ -dimensionnel.

## 5 Résultats

Le modèle générique a été développé à l'aide de classes du langage C++. Pour l'enregistrement des données du modèle, nous utilisons des fichiers au format XML que nous avons appelés « fichiers descripteurs ». La librairie *Xerces-C++* a été utilisée pour lire et écrire dans ces fichiers. Pour la lecture des fichiers issus du plan de traitement (format DICOM-RT) la librairie *DICOM Toolkit* a été utilisée, et des outils ont été développés afin d'intégrer au modèle les informations issues de ce type de fichiers. La méthode de suivi des mouvements a été implémentée en C++ en utilisant les bibliothèques *OpenCV*, et *Coin3D*. Nous avons développé un ensemble de fonctions, mettant en pratique les opérateurs définis dans le modèle générique. Le traitement de données au format VRML associées aux éléments est effectué à l'aide de la bibliothèque *Coin3D*. Nous avons modélisé une salle de traitement de radiothérapie du Centre Léon Bérard à Lyon. La salle est composée du bras d'irradiation (gantry), de la table de traitement (couche), des murs et du plancher.

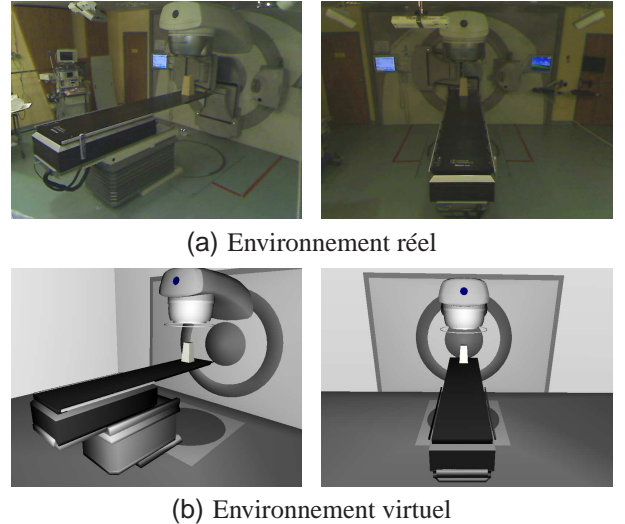


FIGURE 4 – Salle de traitement du Centre Léon Bérard.

Un troisième élément (principal) a été inclus à la composition de la salle pour valider les opérateurs sémantiques : un « fantôme » placé sur la couche de traitement et dont le centre géométrique est situé à l'isocentre (centre du repère global en radiothérapie).

## 5.1 Évaluations

Nous avons commencé par l'évaluation de la méthode sur des séquences d'images générées par ordinateur et de résolution de  $800 \times 600$ . Afin de constituer une séquence d'images qui simule une acquisition « réelle », nous avons fait varier les paramètres des poses des éléments modélisés contenus dans l'ERV de manière aléatoire et nous avons enregistré des images (synchronisées pour l'ensemble de caméras) à différents instants. Les paramètres des poses de l'ERV sont également enregistrés pour chaque image, afin d'obtenir une vérité terrain.

La méthode de suivi a été ensuite testée sur les acquisitions obtenues comme présenté précédemment, afin d'estimer les paramètres des poses des éléments filmés. Enfin, nous avons évalué la différence entre les paramètres simulés (associés à la séquence d'images) et les paramètres calculés avec la méthode. Des exemples de résultats sont présentés sur le tableau de la figure 5.

|              | M      | ET     | EMC     | VSM |
|--------------|--------|--------|---------|-----|
| Gantry (deg) | 0.2620 | 0.6502 | 2.8720  | 5   |
| Couche (deg) | 0.1216 | 0.4174 | 2.0975  | 5   |
| Couche (mm)  | 7.2962 | 6.1743 | 38.5880 | 20  |

FIGURE 5 – Résultats de la précision des calculs sur les acquisitions simulées par ordinateur. M : Moyenne, ET : Écart Type, EMC : Erreur Maximum Calculé, VMS : Valeur Maximum Simulée.

Le graphique de la figure 6, donné à titre d'exemple, montre que malgré de légères fluctuations au niveau local, l'erreur n'est pas cumulative dans le temps. A chaque instant, l'algorithme de calcul de pose converge vers une approximation de la solution cherchée même si un erreur de calcul était présent à l'instant précédent. Le comportement local pourrait être amélioré par l'adoption d'une stratégie de fenêtre glissante qui prenne en compte les poids associés aux poses contenues dans la fenêtre pour calculer la pose courante.

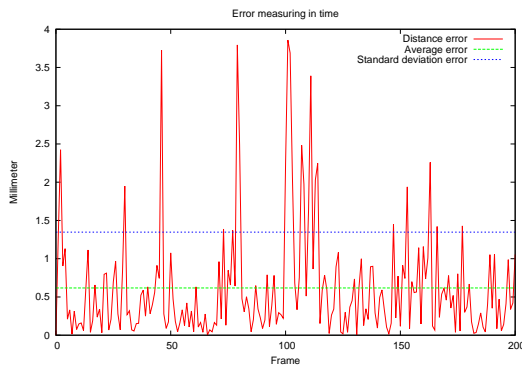


FIGURE 6 – Différences de déplacement pour la partie supérieure de la couche (en millimètres).

## 5.2 Images Vidéo

Après la phase de validation précédente, des acquisitions réelles ont été réalisées. Le système d'acquisition est composée de deux webcams *Logitech QuickCam Sphere*, qui permettent d'obtenir 20 images par seconde à une résolution de  $800 \times 600$ .

La scène et le modèle sont initialisés manuellement en plaçant le bras d'irradiation et la couche en position initiale, dite « position zéro ». À l'instant  $t = 0$  il y a donc une équivalence des poses qui sont évaluées et validées par la phase d'initialisation de la méthode. La phase d'initialisation se termine par l'apprentissage des couleurs de chaque élément et la création d'un modèle d'image du fond, basé sur un modèle de mélange de gaussiennes, pour la détection des mouvements.

Le bras d'irradiation est modélisé comme étant constitué de deux descripteurs « région » et d'un descripteur « point caractéristique ». Ces parties, qui se différencient notamment par la couleur, ont été appelées, respectivement : « bras » (région), « source » (région) et « logo » (point caractéristique). La couche est modélisée en deux sous-parties, cette fois en fonction de leurs degrés de liberté, appelées respectivement : « haut » (région) et « milieu » (région).

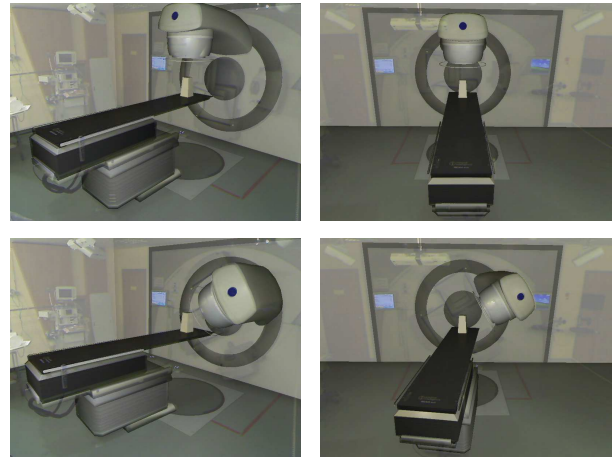


FIGURE 7 – Résultats visuels de l'application (Réalité augmentée).

Dans les images de la figure 7 nous présentons les résultats visuels de la méthode. Nous illustrons l'évaluation d'une pose 3D sur des images réelles segmentées dans la figure 8 pour les deux objets considérés (pour une caméra).

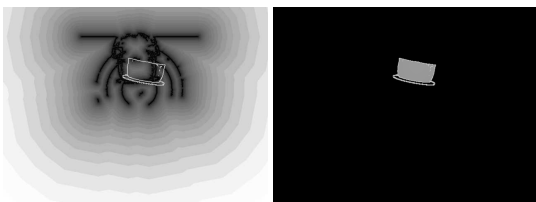
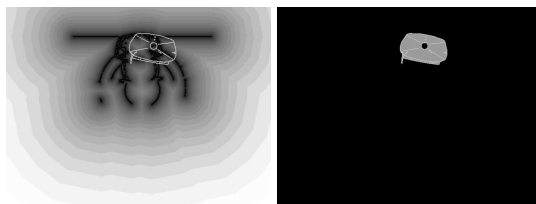
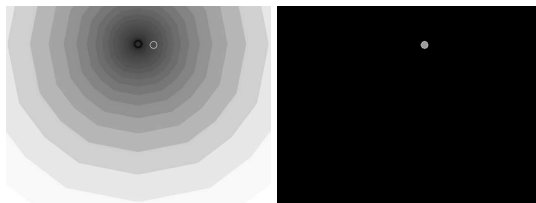
Dans chaque couple d'images (par ligne) :

- l'image de gauche représente la carte de distances aux contours potentiels du descripteur. Les contours obtenus en projetant la partie visible du modèle numérique du descripteur sont superposés sur cette image (avec une pose différente à la pose cible afin de bien illustrer les différences) ;
- l'image de droite contient le masque de visibilité (au niveau de l'image) du descripteur, ainsi que les contours. Pour le bras d'irradiation (Fig. 8(c)), on peut remarquer que

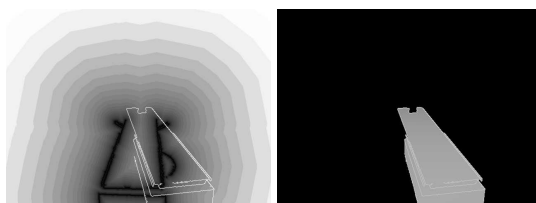
les parties « bras » et « source » présentent plus de bruit, car la couleur de l'objet est proche de celle de l'arrière-plan. Par conséquent, c'est la partie logo qui aidera à lever les ambiguïtés. Cela montre l'importance de la pondération de chaque descripteur.

On remarque que la partie « milieu » de la couche (Fig. 8(e)) n'est pas visible par la caméra. Tous les pixels de la sous-image en bas à gauche des images ont une intensité maximum. Les calculs obtenus sur cette image ne sont donc pas ajoutés à la pose 3D. Ceci illustre la façon dont sont traitées les occultations.

La méthode ne fonctionne pas actuellement en temps interactif. Avec deux objets, le processus génère 2 images par seconde. Ceci pourra être amélioré par la parallélisation de la recherche de la pose et en utilisant la programmation graphique pour générer les projections de descripteurs.



(c) Bras irradiation



(e) Couche

FIGURE 8 – Images générées par notre module.

### 5.3 Le modèle générique

Lors du développement du module on a pu constater l'indépendance du modèle générique par rapport à la méthode de suivi des descripteurs. En effet, les résultats sur la précision du suivi sont plus liés aux performances de la méthode de suivi, cette précision étant cependant proportionnelle au nombre de descripteurs visibles par la caméra car on dispose de plus d'information. De plus, la robustesse du suivi est proportionnelle au nombre de descripteurs associés à l'élément cible. La robustesse est notamment par rapport aux occultations partielles, car un seul descripteur est, par définition, capable de fournir l'information suffisante pour le calcul de la pose. De même, un descripteur peut contenir du bruit (au niveau image) à cause du système d'acquisition ou du changement d'éclairage, et apporter des informations erronées.

Cependant, une description plus fidèle d'un élément permet d'incrémenter les performances de la méthode de suivi. Des contraintes de type « degrés de liberté » et « distance à un autre objet » ont donné de bons résultats, spécialement pour les objets qui doivent rester en contact avec un autre élément.

Le suivi du fantôme inclus dans la salle n'a pas été effectué explicitement, mais, en utilisant le modèle, nous pouvons déduire sa position à partir du déplacement de la couche (et en appliquant la contrainte « est-posé-sur »).

## 6 Conclusion

Nous avons mis en place des outils permettant de créer un environnement de réalité augmentée, à l'aide d'un modèle générique et des fichiers de descripteurs. En utilisant la méthode de suivi présentée, nous sommes prêts à mettre en œuvre un système de contrôle de la radiothérapie qui sera capable de reconnaître des scénarios prédéfinis pour interagir avec la personne réalisant le traitement. Grâce à la collaboration scientifique établie entre DOSIsoft, ETOILE et le Centre anti-cancer Léon Bérard (CLB) à Lyon, nous avons été en mesure de mener les premières expériences présentées dans cet article. Bientôt, nous aurons l'occasion d'effectuer des tests en conditions réelles (après validation des tests dans des conditions équivalentes, mais sans irradiation). Nous travaillons actuellement sur un protocole expérimental avec les radio-physiciens du CLB et les cadres du projet INSPIRA à DOSIsoft.

Les progrès en informatique, physique et robotique conduisent, en particulier, à l'automatisation de traitements médicaux. En conséquence, de nouveaux équipements sont intégrés tous les jours dans salles de soins partout dans le monde, et pas uniquement en radiothérapie. Ces équipements nécessitent la connaissance des positions relatives au patient, comme c'est le cas pour des traitements par radiothérapie. Bien que de tels scénarios soient aujourd'hui plus complexes, notamment par la présence du personnel médical au cours du traitement, nous sommes convaincus que notre système combiné à d'autres pourrait facilement s'étendre à toute situation clinique.

## Remerciements

Cette thèse est financée par une convention CIFRE entre DOSIsoft et le laboratoire LIRIS, dans le cadre du projet INSPIRA, qui implique aussi le Centre Léon Bérard à Lyon, et le centre ETOILE pour la supervision scientifique. Les auteurs tiennent à remercier Mme Chantal Ginestet et Mlle Pauline Dupuis, physiciennes médicales au Centre Léon Bérard, pour leur disponibilité et leur implication, ainsi que MM. Baptiste Germain, Jean-Christophe Diaz, et Hanna Kafrouni à DOSIsoft, et le projet ETOILE.

## Références

- [1] Tobias Sielhorst, Marco Feuerstein, and Nassir Navab. Advanced medical displays : A literature review of augmented reality. *IEEE/OSA Journal of Display Technology ; Special Issue on Medical Displays*, 4(4) :451–467, December 2008.
- [2] E. Samsat, D. Schmalstieg, J. Vander Sloten, A. Freudenthal, J. Declerck, S. Casciaro, Ø. Rideng, and B. Gersak. Augmented reality in surgical procedures. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6806 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, March 2008.
- [3] Nassir Navab, Sandro Michael Heining, and Joerg Traub. Camera Augmented Mobile C-arm (CAMC) : Calibration, Accuracy Study and Clinical Applications. *IEEE Transactions on Medical Imaging*, 29(7) :1412 – 1423, July 2009.
- [4] L.S. Johnson, B.D. Milliken, S.W. Hadley, C.A. Pellizzari, D.J. Haraf, and G.T.Y. Chen. Initial clinical experience with a video-based patient positioning system. *International Journal of Radiation Oncology Biology Physics*, 45(1) :205–213, 1999.
- [5] S. Tao, A. Wu, Y. Wu, Y. Chen, and J. Zhang. Patient set-up in radiotherapy with video-based positioning system. *Clinical Oncology*, 18(4) :363–366, 2006.
- [6] S. Pinault, G. Morel, R. Ferrand, M. Auger, and C. Mabit. Using an external registration system for daily patient repositioning in protontherapy. *IROS 07, IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.
- [7] Stéphane Cotin, Hervé Delingette, and Nicholas Ayache. Real-time elastic deformations of soft tissues for surgery simulation. *IEEE Transactions on Visualization and Computer Graphics*, 5(1), Jan-Mar 1999.
- [8] Frank Tendick, Michael Downes, Tolga Goktekin, Murat Cenk Cavusoglu, David Feygin, Xunlei Wu, Roy Eyal, Mary Hegarty, and Lawrence W. Way. A virtual environment testbed for training laparoscopic surgical skills. *Presence*, 9(3) :236–255, 2000.
- [9] Miguel Portela Sotelo, Elodie Desserée, and Jean-Michel Moreau. Multiple object video tracking for the radiotherapy treatment room supervision : focusing on the model. In *Computer Assisted Radiology and Surgery*. Springer, June 2011.
- [10] E. Marchand and F. Chaumette. Virtual visual servoing : A framework for real-time augmented reality. In G. Drettakis and H.-P. Seidel, editors, *EUROGRAPHICS 2002 Conference Proceeding*, volume 21(3) of *Computer Graphics Forum*, pages 289–298, Saarbrücken, Germany, September 2002.
- [11] Hocine Ouhaddi and Patrick Horain. 3d hand gesture tracking by model registration. In *Proceedings of International Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'99)*, pages 70–73, September 1999.
- [12] Maysam F. Abbod, Diedrich G. von Keyserlingk, Derek A. Linkens, and Mahdi Mahfouf. Survey of utilisation of fuzzy technology in medicine and healthcare. *Fuzzy Sets and Systems*, 120(2) :331 – 349, 2001.
- [13] Cristian Sminchisescu and Bill Triggs. Covariance scaled sampling for monocular 3d body tracking. In *CVPR*, pages 447–454, 2001.
- [14] D.A. Gomez Jauregui and P. Horain. Region-based vs. edge-based registration for 3d motion capture by real time monoscopic vision. In *MIRAGE09*, pages 344–355, 2009.
- [15] Seth Hutchinson, Greg Hager, and Peter Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12 :651–670, 1996.
- [16] Muriel Pressigout, Andrew I. Comport, Eric Marchand, and Francois Chaumette. Real-time markerless tracking for augmented reality : The virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(4) :615–628, 2006.
- [17] A. Yilmaz, O. Javed, and M. Shah. Object tracking : A survey. *Surveys*, 38(4 2006) :13, 2006.
- [18] G. Brogiefors. Hierarchical chamfer matching : A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10 :849–865, 1988.
- [19] Daniel P. Huttenlocher, Gregory A. Klanderman, Gregory A. Kl, and William J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 :850–863, 1993.
- [20] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4) :308–313, January 1965.