Vers une saisie en un seul clic : caractérisation de la forme 3D d'un objet à partir d'informations visuelles.

Towards a one click grasping tool : vision based 3D shape description.

C. Nadeau[1] C. Dune[1]

[1] CEA List, F-92265 Fontenay Aux Roses, France

IRISA, Campus de Beaulieu, 35042 Rennes-cedex
Caroline.Nadeau@irisa.fr *

Résumé

Dans cet article, nous nous intéressons à la saisie d'objets inconnus par un bras robotique équipé d'un système de vision. Suite au travail de [1], une nouvelle méthode permettant d'améliorer l'estimation de la pose et la forme de l'objet à saisir est proposée. Cette méthode s'appuie dans un premier temps sur une reconstruction voxellique de cet objet puis sur l'estimation de ses axes principaux et de ses dimensions. Des choix simplificateurs dans notre implémentation des techniques de reconstruction voxellique assurent un gain en temps de calcul tout en conservant la forme générale de l'objet reconstruit.

Mots Clef

Reconstruction voxellique, Space Carving, saisie autonome, robotique.

Abstract

In this paper, we consider an automatic vision-based grasp of unknown objects. Following [1], a new method is proposed to improve the evaluation of shape and pose of the target. This method is based on volumetric reconstruction of the scene, used to estimate main axis and dimensions of the object to grasp. Our implementation of volumetric reconstruction algorithm is less time-consuming than classical methods, while preserving the object global shape.

Keywords

Volumetric reconstruction, Space Carving, autonomous grasp, robotic.

1 Introduction

Nos travaux s'inscrivent dans un projet de robotique d'assistance à la saisie et à la manipulation d'objets pour les personnes handicapées. L'objectif est de proposer un outil de saisie intuitif et générique, facilement utilisable par des personnes en situation de handicap. Afin de solliciter au minimum l'utilisateur, la saisie est déclenchée par des clics sur une interface graphique, générés au moyen d'un périphérique adapté au handicap (joystick, *smart ball*, commande au souffle, chenillard, etc.).

Dans ce contexte, un robot d'assistance a été développé au CEA-LIST (voir Fig. 1). Le robot SAM (Synthetic Autonomous Majordome) peut évoluer au sein de l'environnement quotidien des personnes en situation de handicap afin de saisir et apporter des objets sur demande. Il est constitué d'un bras MANUS monté sur une plate forme mobile équipée d'outils de navigation et d'une carte de l'environnement. Le système de vision utilisé pour la commande consiste en une caméra déportée sur la base mobile, fournissant une vue globale de la scène et une caméra embarquée sur le bras robotique. Le scénario de saisie est initié lorsque l'utilisateur indique sur la carte d'environnement la station où se trouve l'objet à saisir. Le robot se dirige vers le lieu indiqué et transmet via l'interface graphique la vue de la caméra déportée lorsqu'il atteint sa cible. L'utilisateur désigne l'objet à saisir sur cette vue, déclenchant ainsi la phase de saisie automatique à laquelle nous nous intéressons dans cet article.

Parmi les travaux déjà réalisés autour de la saisie autonome pour le handicap, de nombreuses stratégies consistent à saisir des objets connus ; soit les objets sont marqués, soit leurs modèles ou leurs apparences sont stockés au préalable dans une base de données. Ces approches permettent une saisie robuste d'un ensemble limité d'objets. Cependant, si un défaut de perception du robot provoque une confusion dans la reconnaissance de l'objet observé ou si cet objet est rencontré pour la première fois, la saisie échoue. Si aucune information n'est disponible *a priori* sur l'objet à saisir, comment le détecter dans la scène et quelles sont les informations nécessaires au positionnement de la pince pour parvenir à le saisir ?

*Caroline Nadeau est maintenant affiliée à l'IRISA, UMR 6074.

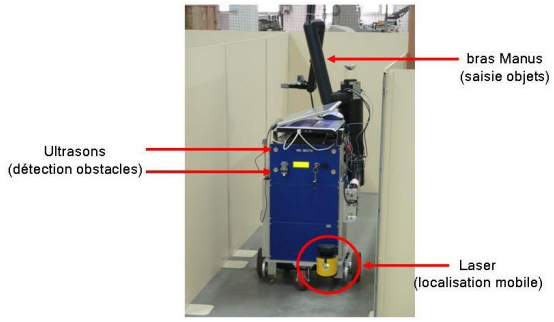

FIG. 1 – Les éléments de la base mobile du robot SAM [7]

Récemment, des travaux se sont portés sur la saisie d'objets inconnus en mileu humain [7, 12, 1, 9, 3, 6]. Pour détecter l'objet à saisir, une solution consiste à solliciter l'utilisateur pour qu'il le désigne dans une vue globale de la scène, par exemple en sélectionnant une boite englobante [7], ou un point [1]. Une fois l'objet détecté, la stratégie de saisie adoptée dépend du type de pince utilisé et de la représentation de l'objet choisie. En limitant la saisie à des objets verticaux et bien séparés, une pince à deux doigts peut être approchée horizontalement vers le centre de la zone désignée [7, 12]. Pour adapter l'approche à une pose quelconque de l'objet, une estimation de son orientation et de sa forme est nécessaire. En supposant que l'objet est globalement convexe, ces données peuvent être grossièrement estimées en calculant les paramètres de sa quadrique englobante, générée à partir de ses contours extraits dans un ensemble de vues [1]. La pince à deux doigts est alors approchée perpendiculairement à l'axe principal de l'objet. L'objet peut également être représenté par un ensemble de primitives géométriques estimées à partir d'un nuage de points obtenu par stéréovision [3]. Enfin, les méthodes de reconstruction voxellique permettent de reconstruire plus ou moins finement tout type d'objets. Dans le cas de la manipulation dextre par une pince multi-doigts [6], un maillage construit sur le volume ainsi obtenu permet de calculer des axes de saisie naturelle pour positionner les différents doigts sur l'objet reconstruit.

Dans cet article, la stratégie de saisie est similaire à [1] et consiste à aligner une pince à deux doigts avec les axes principaux de l'objet. Nous proposons ici une méthode permettant d'améliorer l'estimation de la pose et de la forme de l'objet. A l'approche basée contours de [1] est substituée une approche de reconstruction voxellique qui permet de gérer des objets non convexes et ne nécessite pas d'étape de segmentation des images. Un volume discret, généralement cubique, est positionné virtuellement sur la scène. Chaque élément unitaire le composant, appelé *voxel*, est testé et classé comme appartenant ou non à un élément solide. La reconstruction voxellique permet ainsi de retrouver l'objet discrétisé sans perdre ses concavités contrairement à l'étape d'estimation de la quadrique englobante de [1].

La section 2 de cet article sera consacrée aux méthodes de reconstruction voxellique et aux notions de *photo consistance* et *visibilité* qu'elles mettent en jeu. En section 3 nous détaillerons plus particulièrement l'algorithme de reconstruction voxellique que nous avons développé. Enfin, nous présenterons les résultats obtenus, tant en matière de reconstruction que de caractérisation de la forme de l'objet, en section 4.

2 La reconstruction voxellique

Les méthodes voxelliques fournissent une reconstruction de tous les éléments solides d'une scène à partir d'un ensemble de vues. Dans un premier temps, un volume discret est placé autour de l'objet à saisir, par exemple de manière à englober la totalité d'un support où est posé l'objet. Dans ce cas, ses dimensions sont importantes et un grand nombre de voxels est nécessaire pour obtenir une bonne résolution. Pour accélérer le traitement, un volume de plus petite taille peut également être centré sur l'objet si une estimation de sa position est disponible. Une fois le volume positionné, chaque voxel est testé à l'aide des images de la scène acquises par une caméra calibrée. Ce test porte sur la couleur des pixels sur lesquels chaque voxel se projette dans chaque image. La projection des voxels sur les images est donc une étape essentielle aux méthodes de reconstruction voxellique.

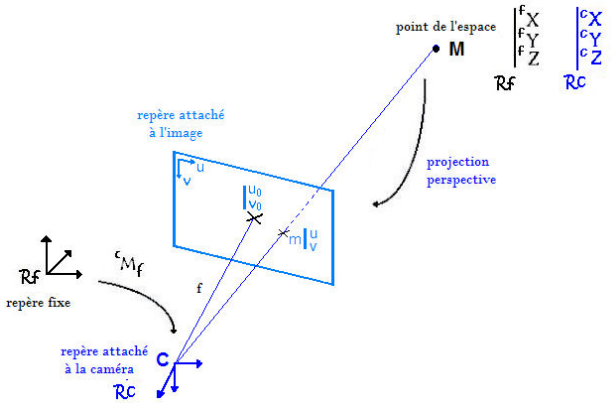

FIG. 2 – Projection d'un point tridimensionnel sur un plan image

Le modèle de caméra utilisé est le modèle sténopé. Pour projeter le voxel V sur l'image de la caméra C (voir Fig. 2), les coordonnées de V exprimées dans le repère fixe sont dans un premier temps transposées dans le repère de la caméra à l'aide de la matrice homogène de changement de repère ${}^{c}\mathbf{M}_{f}$. Le point ${}^{c}M(X, Y, Z)$ est ensuite projeté sur le plan image de la caméra en $m(x, y)$ avec $x = \frac{X}{Z}$ et $y = \frac{Y}{Z}$.

On exprime finalement les coordonnées du point de l'image en pixels $m(u, v)$ à l'aide des paramètres intrinsèques de la caméra.

$$\left\{ \begin{array}{rcl} u & = & u_0 + p_x\, x + \delta_u \\ v & = & v_0 + p_y\, y + \delta_v \end{array} \right. \quad (1)$$

Où δ_u et δ_v représentent les distorsions géométriques du modèle, obtenues lors de la calibration de la caméra [11].

2.1 Notion de photo-consistance

Le test réalisé pour déterminer si un voxel est un élément de l'objet ou du fond est basé sur la notion de *photo-consistance*. Sous l'hypothèse de scènes *Lambertiennes*, un élément de la surface de l'objet sera toujours vu avec la même couleur, indépendamment du point de vue adopté, tandis qu'un élément du fond aura des couleurs différentes sur des images acquises depuis des points de vue différents (voir Fig. 3).

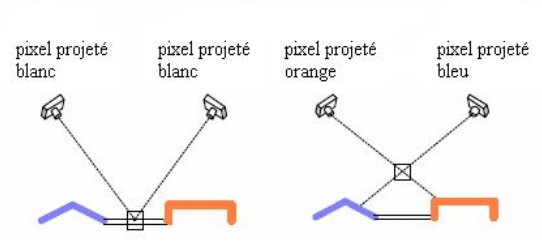

FIG. 3 – Principe de la photo-consistance [8]

En l'absence de bruit, tous les pixels sur lesquels un voxel de l'objet se projette devraient avoir exactement la même couleur. En pratique, pour tenir compte du bruit lié aux réflexions de lumière et à l'acquisition par la caméra, l'écart-type des couleurs des pixels sur lesquels le voxel se projette est calculé. Si cet écart-type est inférieur à un seuil donné, le voxel est déclaré *consistant*, c'est-à-dire qu'il appartient à l'objet. La formule utilisée pour calculer l'écart-type des couleurs est la suivante :

$$\sigma_{V,I} = \sqrt{\frac{1}{K}\sum_{i=1}^{K} I_i^2 - \left(\frac{1}{K}\sum_{i=1}^{K} I_i\right)^2} \qquad (2)$$

Avec K le nombre d'images dans lesquelles le voxel est visible et I_i la couleur du pixel sur lequel il se projette dans l'image i.
Dans le cas d'images couleurs cet écart-type est calculé pour chacun des trois canaux RGB et I_i correspond alors à la valeur entre 0 et 255 du canal testé. Le voxel est considéré comme consistant si l'écart-type est inférieur au seuil pour les trois canaux.

2.2 Visibilité d'un voxel

Un voxel n'est pas forcément visible dans toutes les images acquises, il peut être occulté par d'autres voxels consistants (voir Fig. 4). Pour chaque voxel, seules les images dans lesquelles il est visible sont considérées. Ainsi, avant de tester la photo consistance d'un voxel, sa visibilité est systématiquement évaluée.
Pour tester efficacement la consistance des voxels, le volume initial doit être parcouru de sorte que lorsqu'un voxel est testé, la visibilité de tous les voxels situés devant lui est connue. Cette condition est assurée par un placement

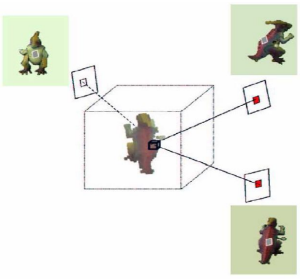

FIG. 4 – Visibilité des voxels d'un objet. Seules les vues de droite où le voxel est visible doivent contribuer au test de photo-consistance [2]

judicieux des caméras qui doivent respecter la *contrainte ordinale de visibilité*. Selon cette contrainte, aucun point de la scène à reconstruire n'est contenu dans l'enveloppe convexe formée par l'ensemble des caméras considérées. Le cas le plus simple consiste à n'utiliser que des caméras situées d'un même côté du volume initial puis de parcourir ce volume plan par plan, par ordre de profondeur croissante.

3 Un algorithme de Space Carving sans test de visibilité

3.1 L'algorithme de Space Carving

L'algorithme de Space Carving [4] autorise un placement arbitraire des caméras. Néanmoins, seules les caméras situées en avant du plan courant et satisfaisant ainsi la *contrainte ordinale de visibilité* sont prises en compte à chaque instant. En général six balayages du volume sont réalisés selon les axes x, y, z dans le sens positif puis négatif de chaque axe.

Tous les voxels du volume initial sont solides. L'implémentation de l'algorithme de Space Carving (voir Algorithme 1) modifie la nature de ces voxels, générant un volume final de voxels transparents et de voxels consistants.

3.2 Test de visibilité des voxels

Un voxel n'est pas visible par une caméra s'il se projette en dehors de l'image de cette caméra ou au même endroit qu'un voxel consistant plus proche que lui du centre de la caméra. En respectant la *contrainte ordinale de visibilité*, lorsqu'un voxel est testé, tous les voxels plus proches que lui du centre de la caméra ont déjà été testés. On peut alors connaître la visibilité d'un voxel V depuis une vue en parcourant la carte de visibilité qui lui est associée. Si le pixel p sur lequel V se projette est répertorié dans cette carte (p est *marqué*), un voxel consistant s'est déjà projeté sur ce pixel et occulte V. Dans le cas contraire, V est visible par la caméra considérée. Par la suite, si V est déclaré consistant, p sera marqué.

Initialisation de l'ensemble des voxels solides
Initialisation du plan courant devant le volume
Faire l'intersection du plan courant avec le volume
Répéter
  Pour chaque voxel V du plan **faire**
    Pour chaque caméra C en avant du plan **faire**
      Calculer la projection p de V sur C
      Si (p non marqué) **Alors**
        Ajouter p à Vis(V)
      Fin Si
    Fin Pour
    Test de consistance de V avec $p \in$ Vis(V)
    Si (V consistant) **Alors**
      Marquer les pixels p de Vis(V)
    Sinon
      V transparent
    Fin Si
  Fin Pour
  Déplacer le plan en profondeur
jusqu'à ce que (Plan derrière le volume)

Algorithme 1: Space Carving

3.3 Limites du test de visibilité

L'efficacité du test de visibilité est liée à la modélisation des voxels adoptée, au rapport entre les résolutions voxellique et pixellique et à la pose relative de la caméra et du volume à reconstruire. Par exemple, en assimilant le voxel à son centre, si la résolution voxellique est faible devant la résolution pixellique, un même voxel se projette sur plusieurs pixels (voir Fig. 5). A l'inverse, si la résolution voxellique est relativement forte, plusieurs voxels se projettent sur un même pixel. Dans le premier cas, des voxels intérieurs sont déclarés visibles et dans le second des voxels de surface sont déclarés occultés.
Pour éviter ces cas d'échec du test de visibilité, des modèles de surfaces dites *étanches* peuvent être construits [5]. Ces surfaces garantissent qu'un rayon reliant le centre optique de la caméra à un voxel intérieur au volume rencontre toujours un voxel de surface. Pour se ramener à un tel modèle de surface, le voxel doit être entièrement projeté dans l'image de la caméra, par exemple en projetant chacun de ses sommets puis en les reliant par une enveloppe convexe [10]. Une empreinte du voxel projeté est obtenue en associant à chaque pixel une valeur de niveau de gris proportionnelle à la surface incluse dans l'enveloppe convexe (voir Fig. 5). L'empreinte ainsi obtenue peut contenir des pixels marqués ou non et doit être analysée pour conclure sur la visibilité du voxel dans l'image.
Une comparaison des modèles de projection ponctuelle et totale des voxels est effectuée dans [10]. La projection de l'empreinte complète du voxel permet d'obtenir un meilleur rendu visuel que la modélisation ponctuelle qui fait apparaître des artefacts. Cependant, dans les deux cas, la forme générale de l'objet est retrouvée et la topologie est conservée.

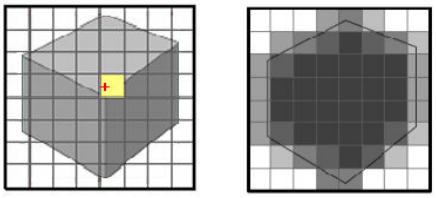

FIG. 5 – A gauche : le voxel est assimilé à son centre et se projette sur un seul pixel. A droite : l'empreinte complète du voxel est considérée [10]

3.4 Suppression du test de visibilité

Si un test de visibilité était implémenté de sorte à garantir l'étanchéité de la surface, il permettrait de reconstruire des objets *Lambertiens* avec une disparité de couleur importante. Cependant la mise en oeuvre d'un tel test est très contraignante dans la mesure où elle nécessite un dimensionnement des résolutions relatives des voxels et des pixels, ce qui fixe la distance entre la caméra et l'objet. D'autre part, elle requiert une projection complète des voxels qui est coûteuse en temps de calcul et apporte principalement une amélioration du rendu visuel.
Notre contribution consiste à proposer un algorithme qui repose uniquement sur la notion de photo-consistance sans tester la visibilité des voxels (voir Algorithme 2). Cette méthode s'apparente à une reconstruction par enveloppe visuelle sans toutefois requérir d'étape de segmentation binaire sur les images acquises. Chaque voxel de l'objet voit sa photo consistance testée à partir des vues d'un ensemble de caméra toujours situé en avant du plan courant traité, qu'il soit visible ou occulté dans chacune de ces vues. L'abandon du test de visibilité limitera notre méthode à la reconstruction d'objets relativement uniformes.
La suppression du test de visibilité et la modélisation ponctuelle retenue entrainent un gain de temps significatif qui permet d'envisager une reconstruction en ligne de l'objet nécessaire à notre application de saise autonome. La section 4 présente les résultats de reconstructions obtenus par un algorithme de Space Carving sans carte de visibilité.

4 Résultats

Pour valider les méthodes de reconstruction voxellique pour une saisie automatique, plusieurs tests ont été réalisés en environnement réel à l'aide d'un robot industriel (voir Fig. 6) équipé d'une webcam.

Dispositif. L'objet est posé sur un support uni ou multicolore et le bras équipé d'une caméra est commandé de manière à obtenir une dizaine de vues de l'objet. Les poses

Initialisation de l'ensemble des voxels solides
Initialisation du plan courant devant le volume
Faire l'intersection du plan courant avec le volume
Répéter
  Pour chaque voxel V du plan **faire**
    Pour chaque caméra C en avant du plan **faire**
      Calculer la projection p de V sur C
    Fin Pour
    Test de consistance de V avec tous les p
    Si (V non consistant) **Alors**
      V transparent
    Fin Si
  Fin Pour
  Déplacer le plan en profondeur
jusqu'à ce que (Plan derrière le volume)

Algorithme 2: Space Carving sans carte de visibilité

de la caméra ainsi que ses paramètres sont conservés pour la reconstruction.

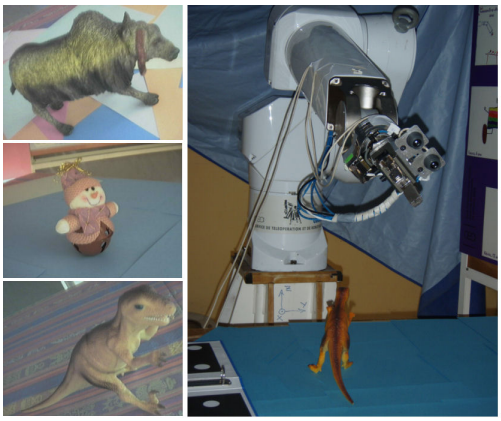

FIG. 6 – Le robot industriel RX90 utilisé pour les essais en environnement réel.

Prétraitement. Dans le cas d'un fond texturé aucun prétraitement des images acquises n'est nécessaire, les pixels n'appartenant ni à l'objet ni au support seront creusés. En revanche, si le fond de la scène est uniforme, un prétraitement est nécessaire pour que le volume soit creusé.
Pour cette étape préalable, nous nous plaçons sous l'hypothèse d'un fond relativement uni dont la couleur correspond à la couleur la plus représentée dans chaque image. Un histogramme de couleurs permet alors de repérer la valeur des composantes RGB de cette couleur dominante et de la remplacer sur chaque image par une couleur déterminée aléatoirement (voir Fig. 7). Cette méthode peu contraignante au niveau des hypothèses de travail (fond uni) permet d'obtenir de bons résultats.

FIG. 7 – Le fond de la scène est remplacé par une couleur aléatoire.

Résultats. La figure 8 illustre la reconstruction d'un objet placé sur un support multicolore. La partie du support incluse dans le volume est conservée en plus de l'objet. Quelques défauts de reconstruction apparaissent (à droite du dinosaure). Ils sont dus au fait que certaines parties du support (feuille orange) n'ont pas une couleur très distincte de l'objet. L'absence de carte de visibilité n'a aucune incidence sur ce défaut qui serait également observé dans le cas d'un algorithme classique. Les points du support ne doivent pas être pris en compte pour calculer les axes principaux de l'objet. Sous l'hypothèse d'un support horizontal, il suffit de parcourir le volume de bas en haut et de supprimer les plans contenant le plus de voxels consistants.

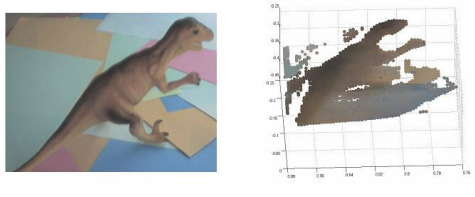

FIG. 8 – A gauche : image acquise avec le robot RX90 de l'objet sur fond multicolore. A droite : reconstruction obtenue sans segmentation préalable du fond.

Sur des supports de couleur relativement unie, la segmentation automatique mise en place permet de supprimer efficacement l'arrière-plan de la scène pour retrouver l'objet. Les reconstructions sont réalisées à partir de vingt vues et avec une résolution voxellique de 2mm. Le temps de calcul de l'algorithme est de l'ordre de 20 secondes. Dans le cas de l'objet "dinosaure" (voir Fig. 9), on obtient une reconstruction assez précise de l'objet avec notamment une bonne reconstruction de ses concavités.

Modélisation. Pour saisir l'objet reconstruit avec une pince à deux doigts, la stratégie adoptée est similaire à celle proposée dans [1]. La position de l'objet et ses axes d'inertie sont estimés par le calcul de ses moments 3D et représentés par l'ellipsoïde correspondant.
La figure 10 présente cette modélisation par un ellipsoïde ainsi que le repère associé à l'objet "dinosaure", reprojeté dans une vue de la scène. Cette représentation tient compte

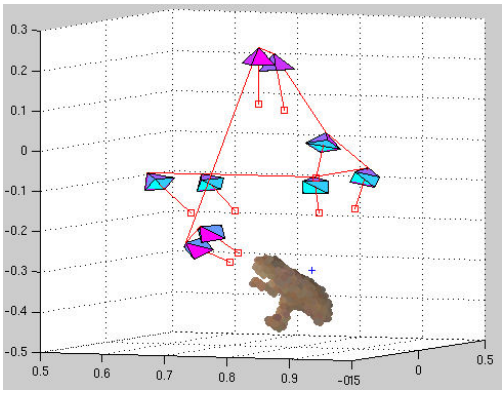

FIG. 9 – Positions des vues de l'objet utilisées pour la reconstruction (1 vue sur 2 affichée) et reconstruction obtenue.

de la répartition des voxels dans le volume en plus de la forme de l'objet. Ainsi, contrairement aux représentations de [1], l'ellipsoïde calculé est essentiellement défini par le corps de l'objet et néglige ses excroissances telles que les pattes de l'objet "dinosaure". Le repère de l'objet est construit à l'aide de son centre et de des axes d'inertie, il est utilisé pour la saisie où le repère lié à la pince lui est aligné.

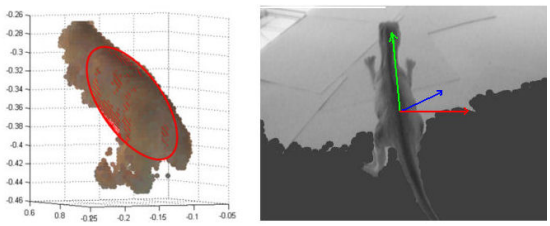

FIG. 10 – A gauche : l'ellipsoïde n'est pas déformé par les excroissances de l'objet. A droite : un repère direct est associé à l'objet reconstruit.

5 Conclusion

Cet article propose une implémentation simplifiée de l'algorithme de Space Carving permettant de reconstruire des objets pour une saisie automatique. En supprimant le *test de visibilité* des voxels et avec une représentation ponctuelle des voxels, l'algorithme est accéléré de manière significative. Il permet de reconstruire des objets de couleur relativement uniforme et distincte de celle du fond. A partir de la représentation voxellique obtenue, il est possible d'attacher un repère ou un ellipsoïde à l'objet pour mettre en place la stratégie de saisie présentée dans [1]. De plus, des informations supplémentaires sont disponibles : certaines relatives aux concavités de l'objet et d'autres aux obstacles qui jonchent la scène. Ces données pourraient être utilisées pour affiner l'approche et la position de saisie.

6 Remerciements

Ce travail a été réalisé au CEA LIST avec le support de la région Bretagne.

Références

[1] C. Dune, E. Marchand, C. Collewet, and C. Leroux. Active rough shape estimation of unknown objects. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'08*, Nice, France, September 2008.

[2] R. Dyer. *Volumetric scene reconstruction from multiple views*, chapter 16, pages 469–489. Kluwer, 2001.

[3] K. Huebner and D. Kragic. Selection of robot pregrasps using box-based shape approximation. In *IROS*, pages 1765–1770, 2008.

[4] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *Int. Journal of Computer Vision*, 38(3) :199–218, July 2000.

[5] C. Leung, B. Appleton, and C. Sun. Embedded voxel colouring. In *Proceedings of Digital Image Computing : Techniques and Applications*, volume 2, pages 623–632, Sydney, 2003.

[6] C. Michel, V. Perdereau, and M. Drouin. An approach to extract natural grasping axes with a real 3d vision system. *Industrial Electronics*, July 2006.

[7] A. Remazeilles, C. Leroux, and G. Chalubert. Sam : a robotic butler for handicapped people. In *IEEE Int. Symp. on Robot and Human Interactive Communication, RO-MAN'08*, Münich, Allemagne, August 2008.

[8] G.G. Slabaugh, W.B. Culbertson, T. Malzbender, and Schafer R.W. A survey of methods for volumetric scene reconstruction. *Int. J. of Computer Vision*, 57 :179–199, 2004.

[9] J. Speth, A. Morales, and P.J. Sanz. Vision-based grasp planning of 3d objects by extending 2d contour based algorithms. In *IROS*, pages 2240–2245, 2008.

[10] E. Steinbach and B. Girod. 3d reconstruction of real-world objects using extended voxels. In *In Proc. ICIP*, pages 823–826, 2000.

[11] R. Tsai and R. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans. on Robotics and Automation*, 5(3) :345–358, June 1989.

[12] K. M. Tsui, H. Yanco, D. Feil-Seifer, and M. Mataric. Survey of domain-specific performance measures in assistive robotic technology. Workshop on Performance Metrics for Intelligence Systems (PerMIS), August 2008.