

Caractérisation de la forme d'un objet  
inconnu et exploration par vision  
active

Le chapitre précédent a présenté une méthode permettant de localiser un *objet générique* à partir d'un *clic* sur l'image acquise par une caméra déportée. Cependant, si aucune hypothèse n'est formulée *a priori* sur la forme et l'orientation spatiale de l'objet, la seule donnée de sa position n'est pas suffisante pour adapter l'approche de la pince et il est alors nécessaire d'obtenir d'autres informations. Pour respecter les critères énoncés dans le premier chapitre de cette thèse ( voir la section 1.1.3), la méthode développée doit permettre de saisir des *objets génériques* (hypothèse 3) positionnés de manière quelconque, dans une scène statique (hypothèse 2)<sup>2</sup>. D'autre part, pour être employée sur un système réel, en milieu humain, l'assistance technique à la saisie que nous proposons doit être robuste aux bruits de mesure et aux imprécisions de la calibration. Enfin, pour être acceptée par l'utilisateur, elle doit permettre, au pire, la saisie en un temps comparable à une saisie en mode téléopéré.

Dans cette thèse, nous nous plaçons dans le cadre d'une approche multi-vues. La littérature traitant de la reconstruction 3D d'objets et de scènes à partir d'informations visuelles est très fournie. Des résultats remarquables ont été obtenus dans le domaine de la reconstruction 3D volumique de scènes statiques ou dynamiques. Par exemple, en détectant la silhouette d'un objet placé sur un fond uniforme (*Visual hull*), par exemple [Franco 06], ou en supposant la scène Lambertienne et en reconstruisant l'objet sous la forme d'un volume discret dont les éléments sont creusés ou conservés selon leur *photo-consistance* [Steven 97, Kutulakos 99]. Même si des tentatives ont été faites pour proposer des versions temps-réel de ces méthodes, par exemple, en proposant des approches pyramidales pour les reconstructions voxelique [Prock 98] ou une parallélisation des calculs pour les méthodes basées silhouette, les méthodes de reconstruction volumique sont assez lourdes à manipuler et gourmandes en temps de calcul. En ce sens, ces méthodes sont bien adaptées pour des applications où les calculs peuvent être faits hors ligne. Le niveau de détail obtenu avec une résolution voxelique importante permet d'envisager la création de personnages pour les jeux vidéos ou encore la reconstruction 3D de sculptures pour sauvegarder une trace du patrimoine conservé dans les musées. Il est cependant légitime de se demander si une reconstruction 3D aussi précise de l'objet est nécessaire à sa saisie, surtout lorsque l'outil de préhension est une pince simple à deux doigts.

Pour saisir un *objet générique* à partir d'informations calculées en ligne, une première stratégie de saisie consiste à utiliser une reconstruction précise de l'objet et calculer ensuite les meilleurs points de saisie physique [Saut 07, Michel 06]. Une autre voie d'étude tente de mimer les prises qu'auraient adopté un humain, par exemple, en repérant sur l'objet l'emplacement de poignées, manches et anses [Edsinger 06, ElKhoury 07]. D'autres approches consistent à approximer la forme de l'objet par un ensemble de primitives géométriques et à calculer les orientations de saisie possibles sur cette forme simplifiée [Huebner 08a, Speth 08]. La motivation de cette thèse est de proposer une méthode de saisie reposant sur le calcul de primitives simples. L'idée sous-jacente est qu'il n'est pas nécessaire de recourir à des méthodes de reconstruction 3D fines pour parvenir à la saisie de la plupart des objets.

De la modélisation de l'objet retenue pour la saisie dépend la méthode d'acquisition des données : pour obtenir une reconstruction exhaustive d'un objet, le système de perception doit être étalonné et considéré comme peu bruité. À l'inverse, si seule la forme générale de l'objet est recherchée, quelques erreurs d'étalonnage peuvent être tolérées. Dans le cadre de l'utilisation d'un système de vision, la stratégie de saisie a également un impact sur le choix des primitives à extraire dans les images : points d'intérêt, contours, régions d'une même couleur.

L'utilisation d'un système de vision pour trouver les informations nécessaires à la saisie induit le problème du positionnement optimal des caméras. En effet, si les positions sont bien choisies, alors les caractéristiques de l'objet sont calculées sans ambiguïtés. À l'inverse, il est possible que les informations calculées soient erronées. En particulier lorsque deux vues sont proches, l'information sur la dimension en profondeur de l'objet est évidemment mal définie. Dans ce cas, l'idéal est d'ajouter une autre vue perpendiculaire aux premières pour ajouter un maximum d'information sur les caractéristiques de l'objet.

---

<sup>2</sup>L'ensemble des hypothèses est rappelé en annexe C

Trois niveaux peuvent être distingués : au plus haut niveau, la stratégie de saisie et le choix de modélisation, puis la perception qui permet la modélisation et enfin, l'extraction des informations dans plusieurs vues dont la position peut être judicieusement choisie.

Dans le premier chapitre de cette partie, la stratégie globale de saisie choisie dans le cadre de notre application sera présentée. Il s'agit d'approximer la forme des objets par des fonctions quadratiques. Le second chapitre rappelle les propriétés de base de ces fonctions paramétriques et le troisième chapitre présente une méthode de reconstruction d'une quadrique à partir de ses projections dans un ensemble de vues. Nous montrerons aussi que le choix des vues influence la reconstruction. En conséquence, le chapitre suivante présentera une méthode de *vision active* permettant de sélectionner les meilleures vues selon un critère relatif à l'information qu'elles contiennent. Finalement des résultats de reconstructions avec un choix des vues par vision active seront présentés. Enfin, au plus bas niveau, le dernier chapitre présente une méthode permettant d'extraire les contours des objets dans des images réelles et les résultats de reconstruction obtenus pour des objets réels.

# Types d'objets et stratégie de saisie

## 5.1 Hypothèses générales

Rappelons que l'objectif pratique soutenant cette étude est d'étendre les capacités de saisie du système Aviso, présenté dans le premier chapitre, à la saisie d'*objets génériques* se trouvant dans une position quelconque. Parmi ces objets, on peut distinguer les objets rigides (tasse, crayon, boîtes, etc.) de ceux qui ne le sont pas (jeux de clés, grappe de raisin, tissu, sachet, etc...). Comme il l'a déjà été spécifié dans le premier chapitre, cette étude traite le cas des objets rigides (hypothèse 4)<sup>1</sup>.

En observant les objets rigides du quotidien, des structures communes qui sont destinées à simplifier leur utilisation peuvent être identifiées : les conditionnements alimentaires sont généralement des contenants parallélépipédiques ou cylindriques (voir la figure 5.1) et les objets manufacturés présentent des axes de symétrie, ou axes naturels de préhension [Michel 06] voire des poignées ou des anses (voir figure 5.2).

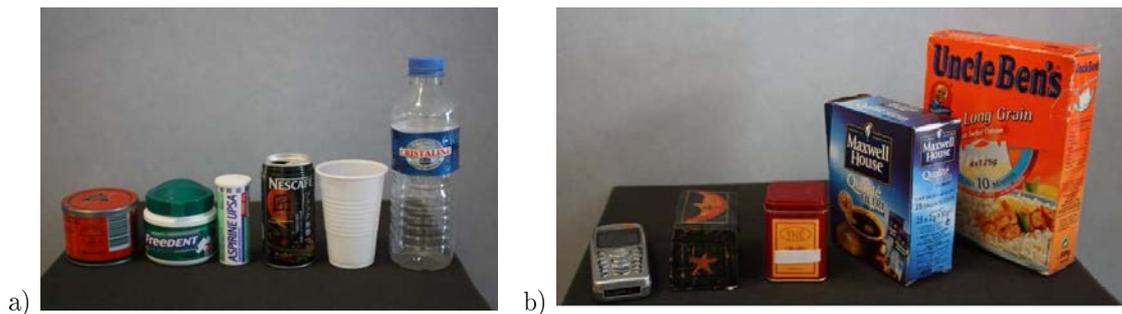


FIG. 5.1 – Objets rigides, globalement convexes et présentant des symétries : a) objets cylindriques (bouteilles, cannettes, conserves, etc.) b) objets parallélépipédiques (boîtes, téléphone, etc.)

Selon le volume des mors de la pince considérée, saisir une tasse par son anse n'est pas toujours possible (voir figure 5.2). D'autre part, ce n'est pas forcément une prise très stable en comparaison aux surfaces latérales de la tasse. Ainsi, une anse n'est pas nécessairement une prise et peut même être considérée comme un obstacle à la saisie. Dans le cadre de cette étude, la saisie sera limitée aux objets

<sup>1</sup>L'ensemble des hypothèses est rappelé en annexe C

rigides qui ne présentent ni anse, ni poignée, ni protubérances. Nous parlerons d'objets *globalement convexes*. À première vue, cette hypothèse semble forte, cependant, elle est vérifiée par un grand nombre d'objets du quotidien (voir figure 5.1).



FIG. 5.2 – Deux objets présentant des prises adaptées à une saisie humaine. Ces prises ne sont pas nécessairement les mieux adaptées à une saisie robotique, en particulier lorsque la pince possède des mors trop volumineux pour être glissés dans l'anse.

**Hypothèse 9** *L'objet est de forme globalement convexe.*

Pour faciliter leur manipulation, les objets conditionnés ou manufacturés présentent des axes de symétrie. Cette hypothèse a déjà été exploitée avec succès pour segmenter un grand nombre d'objets sur un arrière plan quelconque [Li 08] et pour déterminer les axes de saisie d'objets plans [Blake 95, Sanz 05] ou 3D [Hauck 99].

**Hypothèse 10** *L'objet présente un axe de symétrie.*

Comme dans la plupart des travaux de saisie utilisant des capteurs extéroceptifs, les objets sont considérés de densité massique uniforme. D'autre part, comme les objets sont considérés convexes, l'intérieur des contenants sera considéré plein.

**Hypothèse 11** *La densité massique de l'objet est uniforme.*

## 5.2 La stratégie de saisie

En supposant que l'objet à saisir est rigide, globalement convexe, qu'il présente des symétries et est de densité massique uniforme, alors la stratégie de saisie consiste à approcher la pince perpendiculairement à l'axe principal de l'objet en l'alignant avec son plus petit axe. La figure 5.3 illustre cette stratégie de saisie pour quatre objets. Les deux premiers objets respectent les hypothèses fixées et les deux autres présentent des non convexités.

Néanmoins, tant que les excroissances sont suffisamment petites ou placées de façon à ne pas obstruer le chemin de la pince, cette stratégie reste valide. Dans le cas contraire, rien ne garantit son succès et il est nécessaire de détecter ces obstacles et d'en tenir compte pour le positionnement de la pince. La figure 5.4 présente quelques objets non convexes. Dans l'image de gauche, ces trois objets se trouvent dans une pose favorable à notre stratégie de saisie. Aucune convexité non détectable ne fait obstacle à l'approche de la pince. Dans l'image de droite, ces trois objets sont dans une position moins favorable. En particulier, les pattes des animaux-jouets risquent de perturber l'approche de la pince.

Il peut d'ores et déjà être remarqué que cette stratégie engendre plusieurs solutions d'approche de la pince. En effet, la boîte peut être saisie par la droite ou par la gauche, la canette qui est cylindrique peut être saisie depuis n'importe quelle direction, du moment que la pince est perpendiculaire au

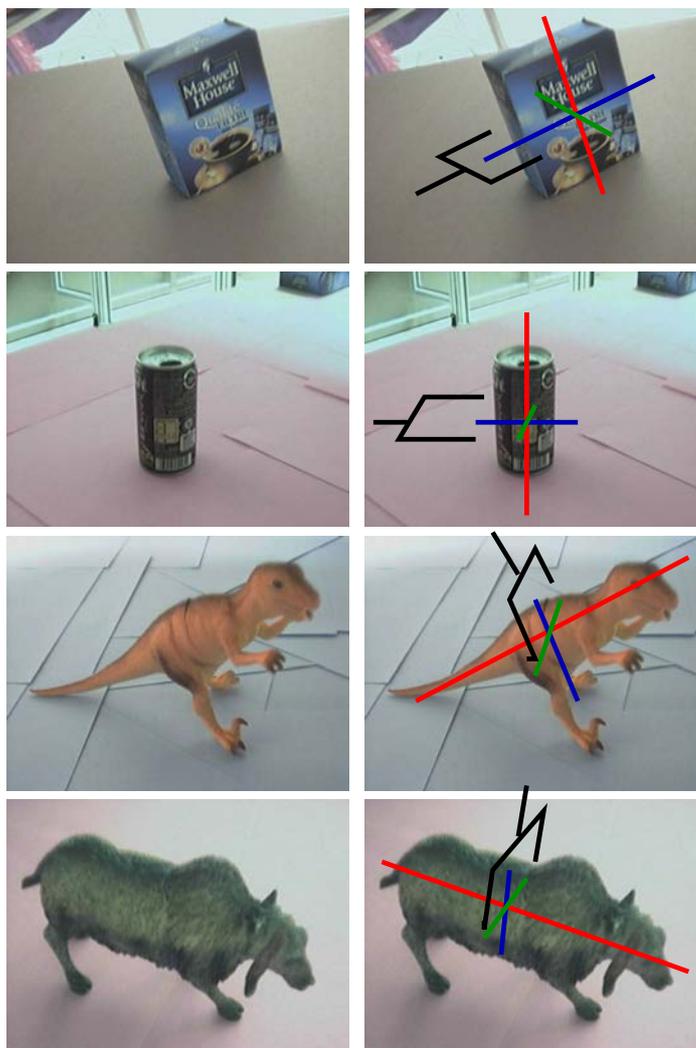


FIG. 5.3 – Stratégie de saisie choisie pour des objets génériques et globalement convexes. La pince est alignée avec les axes principaux de l'objet et centrée sur son centre de gravité. En rouge l'axe principal de la pince, en bleu l'axe moyen et en vert le petit axe.



FIG. 5.4 – Si les objets sont globalement convexes et présentent uniquement quelques concavités suffisamment petites devant les dimensions de la pince, alors ces concavités ne perturbent pas la saisie. À gauche, les objets sont dans une pose favorable à la saisie, à droite, les pattes des jouets risquent de gêner l'approche de la pince.

grand axe. Par ailleurs, dans le cas des deux jouets, parmi les deux solutions de saisie (depuis le haut ou depuis le bas), une seule solution est acceptable puisque l'autre conduirait la pince à heurter la table. Pour choisir l'approche la plus adaptée, il suffit de faire appel à quelques heuristiques simples : en supposant qu'un objet est généralement posé sur un support, seule une saisie "depuis le haut" est possible, si les deux options de saisie respectent cette règle alors la solution choisie sera celle qui demande le moins de déplacement. Si le système dispose d'un module de détection d'obstacle, cette deuxième règle peut être remplacée par : la solution choisie est celle qui permet d'éviter les obstacles.

### 5.3 Modélisation retenue

Les informations nécessaires à la saisie d'un objet sont donc les directions de ses axes principaux et la position de son centre de gravité. Pour obtenir ces informations, la forme de l'objet peut être approximée par une fonction paramétrée de type quadratique. Cette représentation présente quelques propriétés intéressantes pour notre application (voir le chapitre 6) :

- **Une représentation compacte** : une quadrique est représentée par 9 paramètres ;
- **Informations pour la saisie** : les 9 paramètres décrivant une quadrique contiennent l'information de sa position, de son orientation, de sa forme et de ses dimensions, c'est-à-dire directement toutes les informations nécessaires à la saisie ;
- **Projection** : la projection perspective d'une quadrique est une conique. Nous montrerons qu'en mesurant un ensemble de coniques dans un ensemble de vues de l'objet, il est possible de remonter aux paramètres de la quadrique. Ainsi, l'estimation des paramètres de la quadrique correspondant au mieux à la forme de l'objet demande l'extraction des contours de l'objet dans un ensemble d'images.

Notons que ces caractéristiques ont été utilisés par Flandin *et al.* pour décrire les objets composant une scène [Flandin 02]. Ils traduisent les données du centre de l'objet, de son orientation et de ses dimensions en des variables aléatoires qui seront estimées au cours du processus de reconstruction. Le cadre probabiliste leur permet d'inclure les incertitudes de mesure directement dans la modélisation de l'objet et d'obtenir une représentation qui peut être aisément mise à jours à chaque ajout d'information. D'autre part, la représentation choisie permet de sélectionner les meilleures vues par vision active. Un autre avantage de cette méthode est qu'elle permet d'évaluer les caractéristiques de plusieurs objets simultanément.

### 5.4 Acquisition des données

La forme de l'objet peut donc être approximée par une fonction quadratique dont les paramètres sont estimés à partir de la détection des contours dans un ensemble d'images. Il est donc nécessaire de disposer d'un ensemble de vues de l'objet. En utilisant les résultats du chapitre précédent, nous supposons qu'une estimation grossière de la position de l'objet est disponible. Elle est obtenue à partir de la seule information donnée par les coordonnées du *clic* (voir le chapitre 4).

**Hypothèse 12** *Une estimation grossière de la position de l'objet est donnée.*

Les vues de l'objet sont alors acquises sur une sphère de vision centrée sur la position estimée de l'objet. Nous verrons dans le chapitre 9 comment sélectionner ces vues pour améliorer la reconstruction.

Pour extraire les contours des objets dans ces vues, nous supposons que ceux-ci peuvent être segmentés sur le fond à partir de l'information donnée par le *clic* dans la vue globale de la scène. Cette étape de segmentation n'est pas triviale et les systèmes de saisie qui y ont recours imposent des hypothèses fortes sur l'objet et la scène (voir 8.1). Dans le cadre de cette étude, un point de

---

l'objet est connu dans la vue de la caméra déportée et l'information donnée par ce point permet de connaître grossièrement la position de l'objet. Nous montrerons dans la section 8.1 comment exploiter ces informations pour détecter le contour de l'objet.

