Optimisation conjointe du contenu sémantique et du temps d'exécution pour une tâche robotisée de saisie d'objet

Frédéric Lerasle, Joris Guérin, Kimberley Gaume

Optimisation conjointe du contenu sémantique et du temps d'exécution pour une tâche robotisée de saisie d'objet

Kimberley Gaume[1,2], Joris Guérin[1,2], Frédéric Lerasle[1,2]

[1] Université de Toulouse, UPS, 118 route de Narbonne, 31400 Toulouse
[2] CNRS, LAAS, 7 avenue Colonel Roche, 31400 Toulouse

jorisguerin.research@gmail.com

Abstract

Previous work has shown the importance of choosing properly the view under which an object is observed for better identification. Here, we propose to investigate the relationship between execution time and semantic content optimization using a multi-objective formulation. We then perform a grid search to adjust this function's parameter and study the optimization's results on a simulated pick and place example. In practical scenarios, the choice of this parameter will depend on the application context.

Keywords

Semantic view selection, Execution time, Optimization.

1 Introduction

Dans une tâche de *pick and place* sur objet inconnu, la caméra embarquée localise l'objet puis déplace l'effecteur en une pose de prise adéquate pour sa saisie. On privilégie souvent une vue zénithale de la scène qui s'affranchit hélas de toute considération sur la classe et pose courante de l'objet [1, 2]. Il serait opportun de privilégier une vue qui maximise l'information sémantique relative à l'objet, i.e., permettant d'identifier sa classe, afin de robustifier sa perception. Nos travaux antérieurs [3] ont démontré qu'à partir de cette vue zénithale, il est possible d'entraîner un réseau de neurones à prédire la qualité de l'information sémantique dans d'autres vues atteignables par la caméra, et ainsi inférer la pose, souvent différente de la vue zénithale, qui maximisera l'information sémantique de l'image capturée. La pose inférée par ce réseau n'est hélas pas toujours compatible avec la pose de prise à atteindre, ce qui induit un délai dans l'exécution de la tâche de *pick and place*.

Notre approche vise à optimiser la sélection de vue sémantique tout en considérant la pose finale de prise. Le principe est donc d'inférer une pose intermédiaire, entre pose initiale et pose de prise, qui : (i) maximise l'information sémantique dans l'image associée, et (ii) limite les délais induits par cette pose intermédiaire. Nous présentons ci-après notre formalisme ; celui-ci est ensuite validé sur des données simulées.

2 Notre formalisation

On note P la pose du bras manipulateur, caractérisée par une position et une orientation de la caméra fixée sur l'organe terminal. $\{P_i, i \in \{0, ...n-1\}\}$ est un ensemble de n poses du bras tel que la caméra pointe vers l'objet cible. Il représente un échantillon diversifié de poses candidates atteignables par le robot et que l'on doit comparer pour identifier celle qui contient la meilleure information sémantique. P_0 est la pose de départ (vue zénithale), et on note P_n la pose finale du bras à atteindre pour la saisie.

On note $\mathcal{T}$ l'algorithme de planification de trajectoire utilisé, tel que $\mathcal{T}(i,j)$ retourne une trajectoire valide entre P_i et P_j. Étant donnée les vitesses maximales des articulations du robot, on peut calculer le temps d'exécution correspondant, noté $t(\mathcal{T}(i,j))$. Enfin, on note $\mathcal{S}$ la fonction sémantique utilisée, telle que $\mathcal{S}(i,j)$ retourne le score sémantique d'une image qui serait acquise dans la pose P_j uniquement à partir des informations contenues dans une image acquise dans la pose P_i. En pratique, $\mathcal{S}$ est un réseau neuronal convolutif qui prend en entrée l'image correspondant à P_i et la paramétrisation de P_j, et retourne un score de pertinence sémantique entre 0 et 1.

À partir de la pose de départ P_0, nous cherchons à atteindre la pose finale P_n, tout en passant par la pose intermédiaire P_i ($i \in \{0, ..., n-1\}$). L'objectif est de sélectionner P_i de sorte que $F_t(i) = t(\mathcal{T}(0,i)) + t(\mathcal{T}(i,n))$ soit minimal et que $F_s(i) = \mathcal{S}(0,i)$ soit maximal. Il s'agit d'une optimisation multi-critères qui peut être formulée comme suit :

$$\underset{i \in \{0,...n-1\}}{\text{Minimize}} \quad \alpha F_t(i) - (1-\alpha) F_s(i). \tag{1}$$

Avec $0 \leq \alpha \leq 1$ une pondération à définir.

Minimiser F_t ($\alpha = 1$) revient à choisir $i = 0$, ce qui signifie que le robot se déplace de P_0 à P_n sans passer par une pose intermédiaire (utilisant donc l'image obtenue en P_0). Maximiser F_s ($\alpha = 0$) revient à ignorer le temps de trajectoire et choisir la pose la plus informative possible parmi les candidats. Formulé ainsi, à α fixé, le problème peut être résolu par une recherche exhaustive. On souligne l'importance de ramener l'échelle des temps possible à $[0, 1]$ avant d'effectuer l'optimisation.

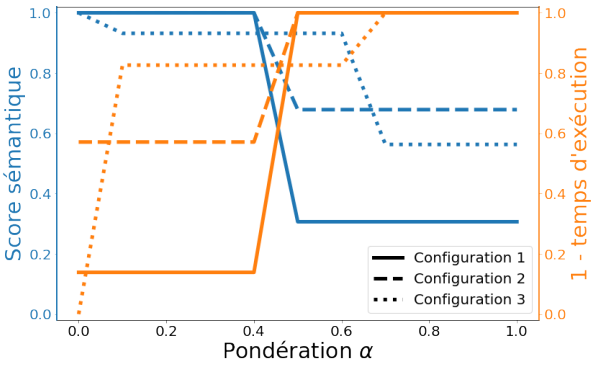

FIGURE 1 – Résultats obtenus pour différentes configurations initiales de l'objet "peigne". Les courbes bleues représentent les scores sémantiques obtenus pour différentes valeurs de α. Les courbes oranges représentent l'opposé du temps d'exécution (1 meilleur temps d'exécution possible, 0 pire temps d'exécution possible).

3 Expérimentations

Une première expérimentation est proposée pour montrer l'influence du paramètre α sur la pose intermédiaire retenue. Nous avons reproduit sur Rviz des scènes correspondant aux conditions d'expérimentation de nos travaux précédents [3]. Il s'agit des trois différentes configurations pour l'objet "peigne" et les poses intermédiaires retenues sont les mêmes que celles présentes dans le jeu de données de sélection de vue sémantique. Les trajectoires du bras robotisé UR10 sont générées grâce à MoveIt, en utilisant l'algorithme RRTConnect [4]. Pour chacune de ces trois configurations, le problème d'optimisation ci-dessus est résolu par une recherche exhaustive pour différentes valeurs de α.

4 Résultats et analyse

Les résultats obtenus pour différentes configurations de l'objet "peigne", et différentes valeurs de α sont présentés en figure 1. On remarque le comportement attendu, pour de faibles valeurs de α, le score sémantique est élevé, mais le temps d'exécution est plus long. À l'inverse, quand α augmente, on s'approche du temps d'exécution optimal en sacrifiant la performance sémantique. On peut également remarquer la présence de paliers sur les courbes, ce qui signifie que les mêmes poses intermédiaires sont choisies sur des plages de valeurs de α. Finalement, cette première expérience montre que la possibilité d'identifier une pose présentant un bon compromis temps d'exécution/score sémantique dépend grandement de la pose dans laquelle se trouve l'objet étudié. En effet, pour la configuration 1, un bon temps d'exécution oblige à avoir un mauvais score sémantique et inversement, alors que pour les deux autres configurations, ce phénomène est beaucoup moins marqué.

5 Conclusion et perspectives

La problématique abordée ici est une extension de nos travaux antérieurs sur la sélection de vue sémantique optimale [3]. Nous proposons une nouvelle formulation qui intègre en plus : (i) la pose finale pré-saisie, et (ii) le temps d'exécution de la tâche associé ; ces considérations sont primordiales pour une tâche de *pick and place* de robotique de manipulation. Notre formulation multi-critère consiste à jouer sur une pondération associée *via* un paramètre α. En pratique, le choix de α va dépendre des contraintes du contexte d'application : si une perte de cadence de production est jugée plus grave que quelques erreurs de tri, on privilégiera des valeurs élevées de α, dans le cas contraire, on privilégiera des valeurs basses.
Nos travaux futurs visent à étendre l'expérimentation préliminaire présentée ici à d'autres objets et intégrer puis évaluer notre système sur un robot réel. Une autre investigation vise à tester l'application dans des contextes où l'impact du passage par une vue intermédiaire sur le temps de trajectoire est plus important qu'il ne l'est ici. Citons par exemple la navigation de robot mobile avec évitement d'obstacles à identifier. Notre stratégie de recherche d'information sémantique optimale est ainsi généralisable à des tâches robotiques autres.

Références

[1] C. Zhihong, Z. Hebin, W. Yanbo, L. Binyan, and L. Yu, "A vision-based robotic grasping system using deep learning for garbage sorting," in *2017 36th Chinese control conference (CCC)*. IEEE, 2017, pp. 11 223–11 226.

[2] J. Guérin, O. Gibaru, S. Thiery, and E. Nyiri, "Cnn features are also great at unsupervised classification," *arXiv preprint arXiv :1707.01700*, 2017.

[3] J. Guérin, O. Gibaru, E. Nyiri, S. Thieryl, and B. Boots, "Semantically meaningful view selection," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1061–1066.

[4] J. J. Kuffner and S. M. LaValle, "Rrt-connect : An efficient approach to single-query path planning," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 2. IEEE, 2000, pp. 995–1001.