

ÉCOLE NATIONALE SUPÉRIEURE DE TECHNIQUES AVANCÉES

UNIVERSITÉ PIERRE ET MARIE CURIE

PROJET DE FIN D'ÉTUDES

Stratégie d'asservissement de robot et de caméra

Auteur

Luigi FRANCO TEDESCO
tedesco@ensta.fr
Promotion 2015

Professeurs Responsables

David FILLIAT
david.filliat@ensta.fr

Safia KEDAD-SIDHOUM
Safia.Kedad-Sidhoum@lip6.fr

Tuteur

Jean-François GOUDOU
jean-francois.goudou@thalesgroup.com

THALES Service | Campus Theresis
828 Boulevard des Maréchaux, 91762 Palaiseau

Stage effectué du 09 mai 2015 au 28 août 2015

Table des matières

1	Introduction	5
1.1	Thématiques du stage.	5
2	État de l'art	6
2.0.1	quelques applications	6
2.0.2	caracteristiques qu'on veut avoir, inspiration humaine	6
2.1	Segmentation	7
2.2	Features	7
2.2.1	Estimation de la normale	7
2.3	Autres features	8
2.3.1	Features 2-Dimensions	8
2.3.2	Features 3-Dimensions	8
3	Description du Travail	9
3.1	Segmentation	9
3.1.1	Algorithme	10
3.1.2	Restrictions	10
3.2	Descripteurs	11
3.2.1	Point Feature Histogram - PFH	11
3.2.2	Fast Point Feature Histogram - FPFH	12
3.2.3	Viewpoint Feature Histogram- VFH	12
3.2.4	Clustered Viewpoint Feature Histogram - CVFH	12
3.3	Classification	12
3.4	Répresentation de l'objet	13
3.4.1	Principes de la Reconnaissance Humaine	13
3.4.2	Charectéristique des objets	13
3.5	Graphe d'aspect polaire	14
3.6	Filtre de Kalman	15
3.7	Chaînes de Markov Chachées	15
3.8	Déplacement du robot	15
3.8.1	Estimation de l'odometrie	16
3.8.2	Problèmes de déplacement	16
3.8.3	Fusion de données	16
3.9	Architecture	16
3.9.1	Movement Primitives	17
3.9.2	Detection and Tracking	17
3.10	Expériment	18
4	Conclusion	19
4.1	Améliorations	19
4.2	Bibliographie	20

Chapitre 1

Introduction

La perception de l'environnement par machines est indispensable pour son intégration à la vie quotidienne. Compétences telle comme se localiser, prise de décisions et capacité d'apprentissage sont nécessaire pour la réalisation de les plus simples tâches. Dans cet étude, on s'intéresse à la compréhension d'éléments constituant d'une scène, sujet récurrent dans le domaine de la vision par ordinateur. Plusieurs approches proposés dans la littérature explorent une sous-partie du pipeline de la reconnaissance pour faire face au difficile défi de représenter la forme visuelle des objets, pendant qu'autres s'intéressent à l'utilisation du système de reconnaissance pour tâches de recherche dans l'environnement, *grasppping*, par exemple. Notre démarche correspond, initialement, à l'intégration de techniques de l'état de l'art pour arriver à un système fonctionnel de reconnaissance incorporé sur une plateforme mobile équipée d'un capteur RGB-D capable. Dans une première temps, la plateforme est capable d'acquérir une base de données d'images d'un objet et, ensuite, d'utiliser ces informations apprises pour vérifier si un objet candidat est ou non celui présenté auparavant, et, finalement, renforcer sa perception avec les information de son déplacement pour lever les possibles ambiguïtés.

1.1 Thématiques du stage.

Vision par ordinateur - Reconnaissance d'objets multi-vue - Robotique

Chapitre 2

État de l'art

2.0.1 quelques applications

La détection d'objets par un robot mobile est traitée dans des articles comme :

Object Search and Localization for an Indoor Mobile Robot [1] et , dont, l'intérêt central, est d'effectuer une recherche/exploration dans un environnement où le plan est connu par avance ce qui permet la discrétisation en zones de recherche, pour retrouver un objet donné a priori. Les systèmes sont capables de retrouver des objets dans des environnements complexes, pourtant cet approche exige un traitement premier pour la localisation.

Learning Search Heuristics for Finding Objects in Structured Environments

Active Object Recognition in Parametric Eigenspace Matrice de covariance entre images à partir de corrélation entre elles. Les vecteurs propres de cette matrice sont utilisées comme base du space. Une image de test est, ensuite, projetée sur cette base et l'image est identifiée avec celle correspondant au plus proche, au sens euclidien, vecteur de la base. Cela apporte aussi une estimation de l'orientation de l'objet, une fois que les images sont labellisées. Pour découvrir quelle est la prochaine action à faire, l'algorithme propose la réduction de l'entropie en s'utilisant des probabilités de reconnaissance à partir des informations antérieures.

Autrement, l'algorithme *Next Best View* traite l'optimisation du nombre d'actions nécessaires par un bras mécanique de six degrés de libertés pour réparer des surfaces chachées et des événements visuels d'un objet à fin d'atteindre un modèle CAD à 3 dimensions. Pour un robot terrestre de déplacement contraint en deux dimensions tout l'information concernant un objet peut être acquis en faisant un tour complet autour du même, ce qui réduit la complexité du espace de recherche. La distance, aussi, joue un rôle important lorsque la résolution des capteurs est limité, mais la capacité de se rapprocher compense ce handicap.

2.0.2 caractéristiques qu'on veut avoir, inspiration humaine

La reconnaissance d'objet est une tâche évident pour des humains et partir dessous sers comme font d'inspiration. Selon les études réalisés en **reference** [5], les humains semblent construire un modèle géométrique rotationnel continu, en autre mots, une composition de vues séquentielles qui font le lien entre différentes faces du objet.

C'est continuité espaciel est, donc, une caractéristique fortement souhaitable. Cette continuité est représentée dans l'article **TODO** où un graphe d'aspect hiérarchique est construit d'après la variation de features.

Autre étude *réf* suggère que l'ensemble de vues augmente le taux de reconnaissance des objet chez les humains, ce qui paraît intuitif.

1. View-based dynamic object recognition based on human perception

Cet article élabore l'importance de une observation temporelle pour la reconnaissance des objets, fait remarqué dans le processus de reconnaissance humaine. La méthode utilisé consisté à partir de l'image initial prise comme key-frame où ses features sont calculées. Ensuite, ses features

sont trackées jusqu'à diminution de features détectées tombe à un nombre inférieur à un seuil définis à priori. La reconnaissance est faite par le comparaison de

4. TableTop Algorithm La segmentation... recognition for each cluster, a simple iterative fitting technique (a distant cousin of ICP) is used to see how well it corresponds to each mesh in our database of models. If a good fit is found, the database id of the model is returned along with the cluster. note that our current fitting method operates in 2D, since based on our assumptions (object resting upright on the known surface of the table) the other 4 dimensions are fixed

5. ENSTA Les travaux qui rasemblent plus l'étude réalisé par cet article sont ceux du [ENSTA](#).... Le premier s'utilise d'un algorithme de segmentation à partir d'extraction de plans correspondant aux murs et au sol, et ensuite classifie l'histogramme PFH global de chaque cluster avec une multi-layer-perceptron.

6. ICUBE Les travaux réalisé par [TODO](#) traité la reconnaissance d'objets multi-vues comme un problème de localisation et suivi par filtre particulaire.

L'intérêt de cet étude est d'avoir une reconnaissance active de l'objet faite par une plateforme mobile.

Répresentation des objets

La figure *xx* dépeint la projection azimutal du modèle décrit. Ce modèle est basé sur

La taille varie decm à ...

Une liste de tous est présent dans les annexes.

2.1 Segmentation

L'algorithme de segmentation tabletop était développé dans le cadre de recherche d'objets sur une table. Les objets sont considérés comme de clusters de points *immédiatement* au dessus du plan de la table, ce qui impose comme contrainte que ce plan doit être le plus important du nuage acquis. La démarche faite par ENSTA, s'utilise du même principe pour réaliser sa segmentation. Le plan plus important est considéré comme sol et les objets sont juste au dessus. De plus, un traitement qui enlève les plans orthogonales à normal du sol aide à l'extraction du fond.

On a choisi un *multi-layer perceptron* comme classificateur. L'entrée du réseaux reçoit les histogrammes de features concaténés un après les autres. Le résultat c'est un histogramme d'entrée de taille *700* - 615 pfh, 100 - sift, 26 -rgb. Le classificateur est entraîné de façon supervisé vers l'algorithme *backpropagation*. La sortie correspond à une couche *softmax* dont sa taille correspond au nombre d'objets entraînés.

2.2 Features

adotamos o pipeline padrao : segmentacao, features e classificacao. (pos-tratamento)

La problématique de la reconnaissance d'objets est largement simplifiée lors qu'une segmentation initiale d'objets candidats peut être faite, cependant, cet étape de selection est un domaine de recherche tout seule. L'avènement de capteurs RGB-D ont donné une nouvelle façade permettant d'exploiter des aspects géométriques pour atteindre la différentiation entre le fond et les objets intéressants. L'algorithme présenté par [ENSTA](#)

2.2.1 Estimation de la normale

Pour constituer les informations géométriques l'estimation de la normale du point est d'extrême importance.

Son calcul est fait de la manière suivante : 1. Un nombre de voisins est choisi 2. Ces points *servent* à trouver des paramètres de l'équation du plan tangent et, par conséquent, la normale correspond.

La méthode adoptée pour la bibliothèque PCL correspond à prendre un certain nombre de plus proches voisins définis par un seuil. Un petit seuil rend le calcul faux et un grand prend en compte points distants qui peuvent ne pas faire partie du plan estimé.

http://www.pointclouds.org/documentation/tutorials/vfh_recognition.php#vfh-recognition

http://pointclouds.org/documentation/tutorials/fpfh_estimation.php

<https://github.com/PointCloudLibrary/pcl/wiki/Overview-and-Comparison-of-Features>

2.3 Autres features

2.3.1 Features 2-Dimensions

—

2.3.2 Features 3-Dimensions

Our stereo data is noisier and sparser than typical line scan data which motivated the use of our new features

- Spherical harmonic invariants [5] : Spherical harmonic invariants and spin images have been successfully used for the problem of object recognition for densely sampled datasets, though their performance seems to degrade for noisier and sparser datasets
- Spin images [6],
- Curvature maps [7] : Conformal factors are based on conformal geometry, which is invariant to isometric transformations, and thus obtains good results on databases of watertight models. Its main drawback is that it can only be applied to manifold meshes which can be problematic in stereoXS
- Point Feature Histograms (PFH) [8]
- Conformal factors [9]

Chapitre 3

Description du Travail

La majorité de la littérature traite le problème de la reconnaissance d'après une seule image de l'objet. Typiquement, un ensemble de *features* est extrait et, ensuite, comparé aux modèles d'objets présents dans une base de données initiale, en contraste aux méthodes directes, comme deep learning, où l'image d'entrée est associée directement avec des classes des objets correspondants au compromis d'une étape d'entraînement importante, pour l'apprentissage de *features*, encore plus dans un espace à 3 dimensions provenant du capteur RGB-D.

Un grand effort était fait pour améliorer l'extraction, le *matching*, ainsi que les *features* elles-mêmes pour qu'elles soient invariantes à transformations affines de l'image et représentatives de l'objet. Ce traitement classique a l'avantage d'être, à la fois modulaire, avec des étapes bien définies de segmentation, extraction de features, classification et pos-traitement, et, au même temps, d'avoir des résultats satisfaisants d'après une implementation plus immédiate. Malgré son intérêt dans certains cas, rapidement on s'aperçoit de limitations lors que vues ambiguës apparaissent.

L'utilisation d'un algorithme de reconnaissance basée sur une seule image apporte l'inconvénient de n'incorporer pas les notions de vue et de transition entre elles, au contraire, la majorité de ces systèmes souhaitent être invariant à les vues d'objets, en autres mots, avoir la capacité de l'identifier de n'importe quel point de vue. Un système dérivé de celui-ci pourrait traiter le concept de vues plus représentatives et transitions, par contre, de façon moins intuitive.

L'objectif ultime c'est d'avoir une reconnaissance multivue, en instance, capable d'incorporer son déplacement pour résoudre des ambiguïtés et faux positifs. Pour incorporer les notions voulus, on présente, simultanément, un simple modèle d'objet suffisamment général et un système capable d'estimer l'orientation de l'objet reconnu, ou bien un système de reconnaissance de vue, pour, ensuite traiter l'information motrice du robot pour augmenter le taux de succès.

3.1 Segmentation

La segmentation consiste de la soustraction des objets d'une image brute, en autres mots, différencier les éléments non constituants du objet de lui-même. La compréhension de la continuité des objets est considérée comme un défi majeur dans le traitement d'image étant donné qu'une fois l'objet séparé du fond, la reconnaissance devient beaucoup plus évidente. Une énorme partie de sa difficulté vient du fait de la projection de la scène dans le plan supprimer l'information correspondant à distance. Les capteurs stéréoscopiques et infra-rouges ont recomposé cette absence d'information et simplifié énormément le traitement nécessaire pour obtenir des objets potentiels. Les cartes de profondeur pourraient être utilisées pour représenter cette nouvelle information, pourtant, encore plus naturelle, le concept de nuage de points propose une représentation spatiale en trois dimensions de l'environnement capturé.

La démarche proposée par la littérature considère les objets comme des ensembles de points défini par un seuil initial de proximité. Cette définition est bien extensif et permet de représenter une énormité, sinon tous, les objets. Néanmoins, définir ces ensembles dans une image brute n'est

pas tout à fait simple. En conséquence, un nouveau à priori qui spécifie que les objets se placent sur des plans de support, malgré plus restrictif que la définition d'avant, permet une segmentation crédible.

3.1.1 Algorithme

La méthode de segmentation du *algorithm Tabletop*, décrit auparavant, se base exactement sur ces aprioris. Pour retrouver les objets posés sur une table, l'algorithme recherche récursivement les plans de support, où le plus important est pris comme la table. Autrement, l'article [ENSTA](#), en partant du même principe, propose un traitement pour le fond de la scène, où les plans orthogonaux à normale du sol et de taille suffisamment grande sont considérés comme des murs, orientés à segmentation d'objets dans les environnements intérieurs. Ainsi, la dernière segmentation, proposée par *Luis Charles*, répond aux exigences du domaine de déplacement du robot : le laboratoire de Thales, Theresis.

Plus spécifiquement, elle peut être découpée dans les étapes suivantes :

1. Soustraction du sol... [DETAILLER](#)
2. Filtrage de points distants, considérés comme plus incertains.
3. Calcul de la normale des surfaces comprises dans la scène
4. Élimination de murs, considérés comme des plans orthogonaux au sol de taille suffisamment grande, d'après un seuil.
5. Projection des points appartenant aux objets dans le plan du sol.
6. Détermination de l'enveloppe convexe correspondant au sol détecté.
7. Réduction de la densité de discrétisation pour accélérer l'étape de *clustering*.
8. Clustering des objets par l'algorithme *point growing*
9. Retour à discrétisation initiale.
10. Calcul du centroïde et *bouding boxes* 2D et 3D
11. extraction d'images, et autres informations pertinentes aux objets détectés.

Une calibration initiale est nécessaire pour définir l'équation du sol. Pour cela, on place le robot dans un endroit de façon que l'image aperçue corresponde majoritairement au sol. L'équation du plan le plus important, le plus grand nombre de points dans le nuage, est extraite par le RANSAC et sauvegardée dans un fichier texte. Une explication plus détaillée sur les sous-méthodes utilisées pour chaque étape, telle que le RANSAC est présentée dans les annexes, ainsi qu'une discussion des paramètres utilisés.

3.1.2 Restrictions

Les physiques des capteurs restreignent les types d'objets qui peuvent être aperçus et, ensuite, segmentés, soit à cause de l'interaction avec les rayons infra-rouges, soit à cause de la résolution limitée des images mesurées. Dans l'autre côté, la segmentation a ses propres contraintes concernant le positionnement des objets dans l'image et, principalement, la définition de sol et murs, résultant dans les restrictions suivantes :

- L'objet se trouve par terre.
- L'objet se trouve au centre de l'image¹
- Environnement isolé de lumière infra-rouge²
- Le sol où le robot se déplace n'est pas accidenté.
- L'objet se trouve à une distance inférieure à 3 mètres³

1. Les objets qui touchent les bords.

2. illumination solaire, par exemple

3. limitation de 5 mètres des capteurs infra-rouges.

- L'objet est assez grand et dépasse le seuil d'appartenance au sol.
- L'objet n'est ni transparents et ni trop réfléchive.

Un grand nombre d'objets, entre chaises tables, écrans, boîtes en carton, poubelles, de tailles et formes variés étaient testés et peuvent être segmentés malgré les restrictions listées. Quelques exemples de segmentation sont présentés dans les annexes pour illustrer la capacité de segmentation.

3.2 Descripteurs

Le travail des descripteurs est, d'un côté, de comprendre les caractéristiques intéressantes et, d'un autre, de réduire la dimensionnalité du espace traité, tandis que restant robuste à des transformations affines et changement de luminosité. On classe les descripteurs selon la caractéristique qu'il exprime. Une première groupe sont les descripteurs géométriques qui essaient de traduire les idées de courbure, forme et taille dans histogrammes, et sont les descripteurs qu'on a donné plus d'attention dans cet étude.

3.2.1 Point Feature Histogram - PFH

Le PFH incorpore les notions de courbure des objets par le calcul de l'écart entre les normales de points. Ce descripteur peut être calculé localement ou globalement, en changeant l'importance du rayon de comparaison. Il est la base d'une grande famille de descripteurs, desquels quelques uns seront expliqués dans la suite.

En revenant à son calcul, l'histogramme est évalué à partir des paires de points à l'intérieur d'un ensemble prédéfini. D'abord, un repère initial, illustré dans l'image *9* est établi sachant le vecteur distance normalisé et les deux normales. Ensuite, trois angles, qui correspondent à la transformation angulaire entre les deux normales, et la distance euclidienne entre les deux points sont estimés. Ces quatre valeurs seront considérées comme features pour réduire le space initial de douze dimension - coordonnées et normales des deux points - à un space de quatre dimension.

$$\begin{aligned} \mathbf{u} &= \mathbf{n}_s \\ \mathbf{v} &= \mathbf{u} \times \frac{(\mathbf{p}_t - \mathbf{p}_s)}{\|\mathbf{p}_t - \mathbf{p}_s\|_2} \\ \mathbf{w} &= \mathbf{u} \times \mathbf{v} \end{aligned}$$

Image d'explication du repère

Puis, les normales sont traduites en features angulaires décrit par les équations :

$$\begin{aligned} \alpha &= \mathbf{v} \cdot \mathbf{n}_t \\ \phi &= \mathbf{u} \cdot \frac{(\mathbf{p}_t - \mathbf{p}_s)}{d} \\ \theta &= \arctan(\mathbf{w} \cdot \mathbf{n}_t, \mathbf{u} \cdot \mathbf{n}_t) \\ d &= \|\mathbf{p}_t - \mathbf{p}_s\|_2 \end{aligned}$$

Le prochain étape c'est de calculer l'histogramme en-soi. Une subdivision du range de valeur de chaque feature angulaire, normalisés pour rester dans le même intervalle trigonométrique, est faite et chaque cellule du histogramme est incrémenté dès qu'une feature tombe dans cet interval.

Le PFH se présente robuste à des différents échelles de densité de points et de bruit, au même temps que invariant à les transformations affines. Des inconvénients viennent de la dépendance de la qualité de l'estimation de la normale⁴.

4. Une discussion des méthodes présentées sur PCL est mis dans les annexes.

3.2.2 Fast Point Feature Histogram - FPFH

L'avènement du FPFH viens de la motivation de réduire la complexité de calcul du descripteur PFH, $O(nk^2)$, pour un nuage avec n points où chaque point a k voisins. Pour cela, l'algorithme au lieu de calculer la relation bidirectionnelle entre tous deux points du ensemble définis, les features de chaque point sont pondérées par les voisins à l'intérieur d'un rayon de recherche, selon la formule au dessous :

$$FPFH(p_q) = SPFH(p_q) + \frac{1}{k} \sum_{i=1}^k \frac{1}{\omega_k} \cdot SPFH(p_k)$$

Ce procedure résult dans une complexité $O(n*k)$. Le gain en vitesse est considerable, ce qui permet des applications en temps réelles. De plus, pour éviter une perte d'information considerable, le FPFH incorpore quelques point externes au rayon de voisinage, mais qui sont compris dans un rayon de taille.

3.2.3 Viewpoint Feature Histogram- VFH

Le VFH, différemment du rapport entre PFH et FPFH, c'est une extension du deuxième descripteur où la variance de point de vue est prise en compte. De forme succincte, des angles entre la normale de chaque point et la direction principale d'observation est concaténé au histogramme provenant du SPFH (Simplified PFH). En gardant le repère utilisé dans les descripteurs d'avant, le vecteur direction principale est défini par la différence entre l'origine du capteur jusqu'au centroïde du *cluster*. Ce résultat permet, au même temps, de reconnaître l'objet et sa orientation spatiale, et, par conséquent, c'est le feature utilisé dans les premiers experiments.

3.2.4 Clustered Viewpoint Feature Histogram - CVFH

CVFH - Clustered VFH - est une feature semi-global capable de gérer occlusions partiels, mauvaise segmentation et bruit par la décomposition du *cluster*, segmenté comme objet, en sous-clusters de structure spatiale homogène. Le descripteur est obtenu d'après un premier filtrage de zones de haute gradient de courbure, considérés comme zones de transitions entre surfaces, et, puis, et l'estimation de l'histogramme VFH pour chaque surface donnée par l'algorithme *point growing*. Ainsi, pour un seul objet, le CVFH ne génère pas un seul histogramme VFH, mais un vecteur des histogrammes. En revanche, le découpage exige un soin un peu plus avec la résolution des surfaces pour lesquelles restent représentatives de l'objet.

3.3 Classification

L'étape de classification correspond à la différenciation entre les histogrammes caractéristiques de chaque vue de chaque objet. Cette mesure pourrait être apprise, par exemple, avec un réseau de neurone ou n'importe quel autre méthode classique de *machine learning*. Dans l'article **Eigen-values object recognition** chaque image est prise comme un vecteur de la base d'un espace euclidien et, donc, la mesure de la distance d'un image de test correspond à la projection de cette image dans les vecteurs de la base. Le travail **6 dof cluster vfh...** suggère l'utilisation de la mesure chi-squared similarité entre histogrammes accolée au classificateur k plus proches voisins. Le grand avantage de ce classificateur c'est l'étape d'apprentissage correspond à la création d'un arbre de recherche, construit d'après la comparaison croisée entre les éléments de la base, que pour l'ordre de grandeur de la base de données envisagée, est presque instantané.

L'API de la librairie FLANN sur PCL permet l'utilisation directe du classificateur K - plus proches voisins. L'implémentation permet l'utilisation de plusieurs définitions de distance entre histogrammes. La définition par défaut, Chi-squared, dont la formule est décrite dans la suite, semble être capable de bien différencier les histogrammes d'entrées et était choisi comme la définition pour le classificateur.

$$\sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

3.4 Représentation de l'objet

Ces choix débouchent sur un système fonctionnel de reconnaissance de vue qui permet de s'intéresser, ensuite, par le couplage de résultat de la reconnaissance avec les informations de déplacement du robot.

3.4.1 Principes de la Reconnaissance Humaine

Commeçnant par le modèle de l'objet, le but c'est de intégrer et respecter certains principes appris après observation dans la reconnaissance chez les humains :

1. Gazltat : Tendance à retrouver des formes et contours simples et naturels par regroupement de caractéristiques et/ou comportements.
2. Continuité : l'apprentissage d'un nouveau objet se fait de forme continue. Dans le cas discret, cela revient à un modèle qui simule les transitions entre superficies.
3. Temporalité et séquentialité : Des études [ref](#) suggèrent que l'ordre de visualization de surfaces des objets influence sa reconnaissance à posteriori. Par conséquent, la sequence spatiale entre vues joue un rôle sur le concept d'objet, où parcourir sequence dans la même ordre que celle appris apporterai plus d'information.

Malheureusement, avoir tous ces principes est une tâche assez complexe pour l'état courant de la technologie, pourtant, en même temps, ils inspirent possibles solutions et représentations. L'apport de cet étude se place dans les domaines de la temporalité et séquentialité.

3.4.2 Charectéristique des objets

En regardant dans la perspective des objet, certaines de ses caractéristiques sont utiles pour le différencier un des autres :

1. Taille
2. Position global
3. couleur et texture
4. Contraintes de space
5. Contexte dans l'environnement
6. Forme géométrique :
 - Sous formes primaire
 - Position et orientation relatif entre formes primaires
7. Affordance : se réfère à le concept d'interactions possibles avec un objet. De manière illustratif, dans le cadre du robot utilisé, cela reviendrait à capacité de pousser un certain objet, d'où l'intérêt de la idéntification de l'orientation du objet.

Le modèle proposé doit être capable d'exprimer au mieux ses caractéristiques en restant, encore, simple. En reprenant la discussion de l'état de l'art, on présent quelques modèles usuellement utilisés pour représenter les objets en trois dimensions.

Modèle CAD

Consiste à représenter l'objet par son modèle 3D fait à l'aide d'outils de design numériques. L'avantage vient du fait d'une fois le modèle construit, la visualisation de l'objet de n'importe quel vue devient évident. De l'autre côté, la fiabilité du modèle est intérieurement liée à la précision de la reconstruction 3D de l'objet, où un soin avec l'échelle et dimensions, ainsi que avec la reproduction de la couleur et texture, est important pour la bonne représentativité.

Évolution de contours

Une autre approche est basée sur les silhouettes des objets et leur évolution d'après transformation affines. Cette problématique c'est démontre mathématiquement compliqué au niveau de la modélisation de fonctions de contour et de leur transformation. Cependant, une fois modélisée, une prévision

Squelettes

...

Aspect-Graphs

Cette forme de représenter les objets consiste à avoir un graphe où chaque noeud correspond à une image d'un point de vue et les liens entre noeuds les réelles transition visuelles. Comme avoir un graphe complet, qui s'approche du continu, apporte un besoin mémoire important et une certaine redondance d'information, la préoccupation principale est de trouver des points de vues représentatives, només *key-Frames*, qui peuvent être choisis avec politiques suivantes :

1. Aléatoire : Ces key-frames peuvent être choisies de forme complètement aléatoire. Absence de calcul intermédiaire ou pre-processing
2. Intervalle constant : Une façon simple c'est de conditionner les *key-frames* à un écart angulaire fixe. Cela permet d'unifier le nombre de frames pour chaque objet, ce qui peut être intéressant pour certaines applications
3. Événement visuels : Cela correspond à déterminer des grandes variations d'intensité des features pour estimer les key-frames plus représentatifs de l'objet. L'inconvénient vient du besoin d'un pré-traitement, en plus, orienté différemment pour chaque feature, lors de la création de la base de données.

3.5 Graphe d'aspect polaire

On considère que les objets sont décrits par deux dimensions d'information : une spatiale, concernant la position absolue de l'objet dans l'environnement et les positions relatives où l'objet était visualisé, et une autre visuelle, donnée par les descripteurs géométriques, de couleurs et de texture ; qu'on cherche à transporter dans un référentiel unique. Le graphe d'aspect permet de coupler l'ensemble d'images suivant ses possibles transitions spatiales ce qui résulte dans la possibilité de construire le modèle à la volée et de jouer avec sa densité d'information - nombre d'images incorporées.

de sorte qu'on a la compréhension des événements visuels.

Formellement, un référentiel polaire entrelace toutes ces informations de façon à représenter la position spatiale d'où l'observation était faite, tel comme il est représenté dans l'image *7*. Pour la construction du modèle les conventions suivantes étaient adoptées :

- l'angle zéro est attribué à la première observation
- L'origine du référentiel est la position globale de l'objet
- Les features sont labellisées d'après le déplacement angulaire et la distance au centroïde de l'objet.

IMAGE

Une grande majorité de features visuelles sont variants à échelle, une fois que la résolution de l'image joue un rôle assez critique pour la détection de features, comme les patch SIFTs. Ainsi, avoir la distance que l'image était prise peut être intéressant pour limiter la classification à une échelle valable.

3.6 Filtre de Kalman

La modélisation des objets entraîne le besoin initial de les localiser dans la scène pour, postérieurement, les identifier. À cause de la divergence de l'odométrie, la mauvaise segmentation et le calcul du centroïde de l'objet, la position estimée est fortement bruitée ayant un écart type qui rend la suite et l'identification infaisable lorsque plusieurs objets sont minimalement proches. Un filtre de Kalman ayant un modèle unitaire pour la matrice de transition d'états, moyenne les observations pour s'adapter au bruit de mesure.

Cependant, le caractère mono-modal du filtre de Kalman fait en sorte qu'un seul objet cible peut être suivi à la fois. Pour atteindre l'aspect multi-modal, il faut que plusieurs filtres tournent en parallèle. Ainsi, le problème passe de estimer la position à décider quel observation appartient à quel filtre, l'étape d'identification. Cela se fait à l'aide d'une matrice de corrélation de distances entre les nouvelles observations et les états courants de chaque filtre existant. Une solution simplificatrice est d'associer chaque observation au filtre selon l'ordre de vraisemblance de cette matrice. Lors que une ambiguïté se produit dans l'étape d'identification, la classification peut aider à prendre une décision de mettre un filtre à jour ou, alors, créer un nouveau filtre.

3.7 Chaînes de Markov Cachées

Le déplacement physique du robot résulte dans une séquence d'observations, en angles différents, d'un même objet. On exploite l'information odométrique entre les visualisations pour prédire les prochaines possibles orientations. De cette manière, l'évolution de la reconnaissance au long du temps est représenté par un processus stochastique, dont une modélisation possible correspond à le traiter de façon discrète dans un espace d'état. Ayant l'appriori que la dernière image et le dernier déplacement suffisent pour faire cette prédiction, en respectant, donc, la propriété de Markov de premier ordre, le processus stochastique est modélisée sur le cadre d'une chaîne de Markov cachée.

Concrètement, les états cachés correspondent à des objets connus au préalable et déjà incorporés dans la mémoire du robot. Cela contraint le nombre d'états et on se rencontre avec un chaîne fini. Puis, une matrice de transition décrit l'évolution du processus et c'est là où l'odométrie et la relation entre vues et entre objets sont incorporés. Finalement, une autre matrice, dit matrice d'émission, estime la vraisemblance entre l'observation et les états de la chaîne.

Une autre deuxième modélisation serait d'avoir une chaîne de Markov Cachée distincte pour chaque objet et ensuite décider quel était le processus le plus vraisemblable. Ce cas est un sous-ensemble du cas antérieur où les transitions entre deux objets ne sont pas considérés. Pourtant, ce qui peut être utile s'on considéré l'évolution d'objets, par exemple, la transition entre une chaise vide et une personne assise sur une chaise ou encore un personne commence à marcher.

3.8 Déplacement du robot

Le robot est équipé de trois roues, desquelles les symétriques arrières sont motorisés et responsables pour le déplacement motrice. Au même temps que la dernière sert à donner un support pour la partie derrière de la carrosserie. Les moteurs sont contrôlés à partir de commandes séries, pré-établis pour le fabricant, qui définissent la vitesse de roulement. La combinaison de les rotations des deux roues motorisées dans les deux sens possibles permet au robot d'avoir les comportements suivants :

- Déplacement en ligne droite : équivalent à les deux roues roulant avec la même vitesse et dans le même sens.
- Déplacement en arc de cercle : La différence entre les vitesses de roues résulte dans un mouvement de cercle. Le rapport entre cette différence permet de définir la courbure de la trajectoire.
- Rotation : Dans ce cas, les deux roues sont commandées à la même vitesse, mais avec des sens différents.

Illustration

Finalement, la combinaison de ces mouvements permet au robot d'accéder à n'importe quelle position du space.

3.8.1 Estimation de l'odometrie

Certains robots sont dotés de capteurs aptes d'estimer de façon approximative son déplacement. C'est aussi le cas du robot ciblé qui possède des encodeurs capables d'estimer la rotation angulaire des roues. Une intégration, au sens mathématique, de la différence entre l'odometrie entre deux intervalles de temps permet de retrouver la position globale du robot.

* FORMULE *

3.8.2 Problèmes de déplacement

La roulette de support originalement installée avait deux axes de rotation. Pourtant, quelques mouvements de rotation du robot alignent la roulette orthogonalement au sens du prochain mouvement ce qui crée une trajectoire parasite qui perturbe la trajectoire voulue. Une tentative frustrée d'installer une bille omnidirectionnelle à roulement, qui se bloquait sur la moquette avec le poids du robot, a fait que l'originale était réinstallée. Une deuxième solution serait d'interdire certains mouvements du robot pour éviter cette déviation.

3.8.3 Fusion de données

L'estimation de l'odometrie diverge au long du temps dû à l'accumulation d'erreurs de mesure. Cette divergence est encore plus considérable. Dans l'autre côté, l'utilisation du capteur RGB-D estime la distance au centre de chaque objet. Une correspondance entre les objets de deux observations consécutives nous donne une autre référence de positionnement. Un couplage des deux mesures, une provenant du capteur moteur et l'autre du capteur infrarouge, fait que l'odometrie doit être

3.9 Architecture

Le design de l'architecture sert à décider comment définir les unités de traitement et la communication entre elles. La définition des unités de traitement suit le découpage du pipeline de reconnaissance avec des nœuds responsables pour la segmentation, calcul de features et la classification, aussi comme, le contrôle du robot.

L'interface matériel-logiciel était faite sur l'environnement ROS - Robot Operating System. Aussi que les outils d'affichage, ROS, rassemble les bibliothèques d'acquisition d'images RGB-D, OpenNI 2 et Freenect, et de traitement de nuage de points, PCL. De même, sa structure de nœuds a permis une implémentation modulaire et directe du système décrit au-dessus, bien comme la communication entre machines, l'ordinateur portable et celui embarqué au robot.

L'image *10* illustre l'architecture en soulignant le flux d'information à travers des nœuds.

3.9.1 Movement Primitives

3.9.2 Detection and Tracking

L'idée initiale était d'utiliser une caméra pan-tilt-zoom (ptz) comme capteur principale, pourtant,... n'est pas explorer des information en plus donnés par les capteurs ...

In other to follow the object, a RGBD [Microsoft Kinect] is used. This device provide depth information about the image, therefore, increasing the amount of information we receive. Treating point cloud instead of pixels made it possible to apply a direct treatment to remove the unnecessary parts of the image such as walls and the floor. This segmentation occurs in the following way :

3.10 Expériment

D'abord, vingt objets de tailles et formes diverses ont été choisi pour évaluer les capacités *réognitif* du robot. Ils sont objets typiques qui peuvent être facilement retrouvés dans un laboratoire. Une liste avec tous est incorporé aux annexes. Ensuite, le feature VFH était calculé pour huit positions différents écartées de 45 degrés. La position correspondant au angle zero, était choisi de manière aléatoire en alignant un des axes de l'objet avec celui du capteur.

Une première évaluation proposée consiste à faire un tour complet autour de l'objet à être reconnu en quatre positions angulaires différentes : 0, 45, 90 et une dernière choisi de manière aléatoire pour chaque objet. Le robot fait le tour à une vitesse de $0.35 \pm 0.1m/s$ à une distance de $1.5m$, en enregistrant des images à $1hz$, ainsi, une expérience typique consiste d'environ 25 images d'angle différent et prendre 25 ± 3 seconds. Ensuite, trois différentes ratios sont calculés pour exprimer la reconnaissance d'objets, la reconnaissance de vue et la suivi des reconnaissances par la chaîne de Markov cachée.

Dans le premier tableaux on retrouve le résultats de la reconnaissance donné par la comparaison des histogrammes provenant du *plus proche voisin*. Ce résultat estime la capacité de distinguer deux objets quelconques, en autre mots, cette capacité viens de la représentativité des descripteurs utilisés et l'efficacité de la mesure de similarité entre histogrammes.

Chapitre 4

Conclusion

L'apport de ce projet *relie* se place dans la partie finale du traitement/pipeline de la reconnaissance. Au remarque que le pos-traitement proposé pourrait être fait pour n'importe quel système de reconnaissance avec estimation de l'orientation du objet reconnu et un système mobile capable d'informer son déplacement. La reconnaissance d'objet multi-vue augment la capacité de résoudre situations d'ambiguïté et gère les problèmes provenant de l'absence de vues base de données et erreurs de segmentation, en conséquence ...

4.1 Améliorations

...

4.2 Bibliographie

[1] Active Object Recognition in Parametric Eigenspace

[2] View-based dynamic object recognition based on human perception

[3] L.-C. Caron, D. Filliat, A. Gepperth. Neural Network Fusion of Color, Depth and Location for Object Instance Recognition on a Mobile Robot. Second Workshop on Assistive Computer Vision and Robotics (ACVR), in conjunction with European Conference on Computer Vision, Sep 2014, Zurich, Switzerland.

[4] Radu Bogdan Rusu, Gary Bradski, Romain Thibaux, John Hsu, Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram

[5] G. Burel and H. Henocq, Three-dimensional invariants and their application to object recognition," *Signal Process.*, vol. 45, no. 1, pp. 1–22, 1995.

[6] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1999.

[7] T. Gatzke, C. Grimm, M. Garland, and S. Zelinka, "Curvature Maps for Local Shape Comparison," in *SMI '05 : Proceedings of the International Conference on Shape Modeling and Applications 2005 (SMI' 05)*, 2005, pp. 246–255.

[8] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in *ICRA*, 2009.

[9] B.-C. M. and G. C., "Characterizing shape using conformal factors," in *Eurographics Workshop on 3D Object Retrieval*, 2008.

[10] K. Lai, L., X. Ren, & D. Fox A Large-Scale Hierarchical Multi-View RGB-D Object Dataset In *IEEE International Conference on Robotics and Automation (ICRA)*, May 2011.

[11] R. B. Rusu and S. Cousins, "3D is here : Point Cloud Library (PCL)," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Shanghai, China, May 9–13, 2011, pp. 1–4.

[12] Three Dimensional object recognition and 6 dof pose estimation

Sites internet

http://pointclouds.org/documentation/tutorials/fpfh_estimation.php

<https://github.com/PointCloudLibrary/pcl/wiki/Overview-and-Comparison-of-Features>

http://www.pointclouds.org/documentation/tutorials/vfh_recognition.php#vfh-recognition

http://robotica.unileon.es/mediawiki/index.php/PCL/OpenNI_tutorial_4:_3D_object_recognition