

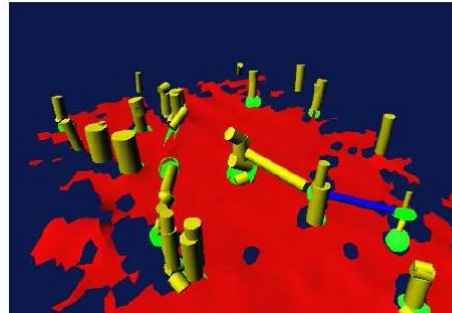
ASTRID: Analyse de Scène en Temps Réel pour l'Interaction 3D

Directeur : Tamy Boubekeur, Telecom ParisTech

Encadrant : Jose Alonso Ybáñez Zepeda, Ayotle

Résumé : L'objectif de cette thèse est de créer automatiquement et exploiter une carte intelligente de l'environnement 3D réel d'un utilisateur, qui permette de le rendre interactif afin de contrôler à la fois des entités réelles (objets connectés) et virtuelles. On souhaite ainsi développer de nouvelles méthodes d'analyse de scène 3D non structurée et inconnue a priori, afin d'en obtenir une interprétation haut niveau offrant au système hôte une perception des objets la constituant. L'état de ces objets permettra de définir diverses zones d'interaction accessibles aux utilisateurs présents dans la scène. L'interprétation passera par une modélisation géométrique qui a pour but de détecter et représenter par des primitives géométriques 3D les éléments de la scène en analysant les formes. Une classification et une analyse sémantique de ces formes pourra compléter la modélisation de la scène dans la seconde partie de la thèse. Les mouvements des objets devront alors être détectés et analysés pour créer de nouvelles zones d'interaction. Le modèle de scène en entrée prend typiquement la forme d'un nuage de points 3D dense et non structuré qui sera acquis et mis à jour à l'aide d'un système embarqué de numérisation 3D de type *camera RGB-Depth* temps réel. Ce dernier pouvant lui même être en déplacement, on considérera les problématiques de localisation du système dans l'environnement et de contrainte temps réel sur des appareils embarqués. L'objectif scientifique principal de la thèse est de développer une méthodologie d'analyse de forme temps réel qui soit robuste à ce scénario, en exploitant les avancées récentes en traitement géométrique, reconnaissance et classification de forme et interaction computationnelle.

Mots clés : *interaction 3D, vision par ordinateur, analyse de scène, interprétation de scène, géométrie numérique, modélisation géométrique, analyse sémantique, classification, cartographie 3D, localisation 3D, systèmes embarqués, temps réel, réalité augmentée, objets connectés*



Les techniques de modélisation géométrique de scène permettent une interprétation sémantique de la scène qui peut être utilisée par exemple dans le cadre de systèmes de navigation autonomes embarqués. L'objectif de cette thèse est d'utiliser cette interprétation pour permettre l'interaction de personnes avec leur environnement.

Contexte

Cette thèse de doctorat s'inscrit dans le thème de l'*analyse haut niveau* de scènes tri dimensionnelles.

La recherche du mode idéal de perception d'une scène inconnue par un système de calcul a connu de nombreuses avancées depuis plusieurs années et les études récentes menées dans le domaine de l'intelligence artificielle, et plus précisément des systèmes de vision, ont permis de faire rendre ces méthodes d'analyse compatibles, dans une certaine mesure, avec le traitement de nuages de points 3D denses et imparfaits issus de la capture (bruit, points aberrants, zones manquantes, occlusions).

Les systèmes actuels d'acquisition de scène permettent d'obtenir un nuage de points dense représentant précisément l'environnement. Dans le cadre du scénario présenté ci-dessus, nous travaillerons avec des caméras 3D de type Time-of-flight, Light coding ou Stéréovision. Bien que moins précises que les LIDAR haute-performance, ces caméras grand public sont plus accessibles.

C'est à partir de ce type d'images tridimensionnelles que la modélisation de la scène peut être faite. En effet, en réalisant un traitement géométrique pour analyser le nuage de points, il sera possible de mettre en lumière les orientations des ensembles de points.

Des techniques comme la transformée de Hough, permettant de détecter les alignements de points, ou encore l'analyse en composantes principales, permettront de dégager des groupes de points constituant des formes distinctes.

On pourra ensuite utiliser des descripteurs 3D (ou features) pour détecter les coins et contours des formes 3D. Ces points et contours caractéristiques permettront de segmenter précisément les formes représentant les différents objets de la scène.

L'étape suivante reviendra à associer à chaque objet segmenté, une primitive géométrique telle qu'un cube, parallélépipède, cylindre, cône, ellipsoïde, etc, qui représente de la manière la plus fidèle la forme d'origine de chaque objet.

La capacité de calcul des ordinateurs est aujourd'hui en évolution constante, cependant les besoins en calcul de l'analyse d'un nuage de points dense ont une importance qui reste à prendre en compte dans l'objectif du traitement en temps réel.

Description du travail

L'activité de recherche menée durant la thèse s'articulera autour de 5 défis applicatifs principaux :

1. Repositionnement des zones interactives suite à des mouvements légers de la position du capteur 3D (déplacements de quelques centimètres autour d'une position avec une rotation de moins de 10°).
2. Repositionnement des zones interactives suite à des changements importants de la position du capteur 3D (déplacements de quelques mètres, avec des fortes rotations).
3. Modélisation de la scène 3D (détection d'objets).
4. Visualisation simplifiée pour la conception des zones interactives.
5. Reconstruction de la cartographie.

Pour atteindre ces objectifs, dans le cadre d'une exécution temps réel, on développera un travail méthodologique organisé en 3 temps.

Tout d'abord, on établira un état de l'art précis dans le domaine de la reconnaissance de formes 3D temps réel, que l'on agrémentera des éléments pertinents liés aux méthodes de consolidation robuste [Bouaziz13] [Mellado14], aux transformées dans l'espace des paramètres [Decoret03] et à la co-analyse [Averkiou15].

Dans un second temps, on mettra en oeuvre un système complet de repositionnement des zones interactives combinant détection de singularités (descripteurs ou features), transformée dans l'espace des paramètres et éventuellement apprentissage statistique. Ce dernier aspect visera à exploiter au maximum les alignements précédemment effectués pour une zone interactive, et leur corrélation aux données capturées (comparaison des formes des zones et sous-groupes du nouveau nuage de points), pour aider et accélérer le plus possible l'alignement courant. On envisage également d'étudier l'approximation de formes [Thiery13] et, de la même manière que les *super-pixels* permettent d'interpréter efficacement des vidéos en temps réel, de développer le concept de *super-points*, en combinant simplification géométrique et détection de features pour réduire la quantité de données 3D à analyser. La plupart des zones interactives ayant le volume et la structure de primitives géométriques simples (rectangles, cones, sphères, cylindres), on développera une méthodologie basée sur la transformée de Hough, avec une implémentation temps réel, pour effectuer le recalage global lors de grands mouvements.

Dans un troisième temps, on s'intéressera au design des zones interactives et à la construction par l'utilisateur d'une cartographie complète et interactive de son environnement s'appuyant sur une reconstruction de la scène. Il s'agit de la phase la plus exploratoire de la thèse où le challenge sera d'offrir des mécanismes simples, intuitifs et guidés par les données pour construire ces cartographies, en éliminant le besoin express d'un passage des données par des outils experts type CAO. On pourra s'inspirer de méthodes de modélisation de scène haut-niveau [Lalonde06], basées sur la détection de descripteurs 3D et de l'orientation locale du nuage de points, pour ajuster des formes géométriques simples à des objets détectés dans la scène capturée et en faire des zones interactives.

Environnement de Travail

Cette thèse sera effectuée à la fois chez Ayotle, encadrée par Alonso Ybanez (docteur en vision par ordinateur) et à Telecom ParisTech, sous la direction de Tamy Boubekour, au sein de l'équipe de recherche qu'il dirige (équipe d'Informatique Graphique 3D, Groupe TII, Département TSI, UMR CNRS 5141, Telecom ParisTech, Université Paris-Saclay). Les travaux récents de l'équipe pourront alimenter la réflexion en début de thèse. La thèse bénéficiera des collaborations académiques en cours de l'équipe sur les sujets connexes, avec notamment TU Vienna, TU Delft, CNR Pisa, TU Berlin, New York University, Osaka Université, Dassault Systèmes, Ubisoft et Disney Research entre autres.

Le travail de thèse sera régulièrement expérimenté en conditions réelles, via des démonstrateurs implémentés pour une plateforme embarquée de type Raspberry Pi2 (OS Linux Raspbian, Camera 3D Asus Xtion Pro).

Références

[Decoret03] Billboard clouds for extreme model simplification.

Xavier Décoret, Frédo Durand, François X. Sillion, and Julie Dorsey.

SIGGRAPH 2003

[Lalonde06] Natural terrain classification using three-dimensional lidar data for ground robot mobility

JF Lalonde, N Vandapel, DF Huber et M Hebert.

[Li11] GlobFit: Consistently Fitting Primitives by Discovering Global Relations

Yangyan Li, Xiaokun Wu, Yiorgos Chrysanthou, Andrei Sharf, Daniel Cohen-Or, Niloy J. Mitra

ACM SIGGRAPH 2011

[Chen 13] 3-Sweep: Extracting Editable Objects from a Single Photo

Tao Chen, Zhe Zhu, Ariel Shamir, Shi-Min Hu, Daniel Cohen-Or

SIGGRAPH 2013

[Thiery13] Sphere-Meshes: Shape Approximation using Spherical Quadric Error Metrics

Jean-Marc Thiery, Emilie Guy and Tamy Boubekeur

SIGGRAPH Asia 2013

[Bouaziz13] Sparse Iterative Closest Point

Sofien Bouaziz, Andrea Tagliasacchi, Mark Pauly

Symposium on Geometry Processing 2013

[Mellado14] Super4PCS: Fast Global Pointcloud Registration via Smart Indexing.

Nicolas Mellado, Dror Aiger, Niloy J. Mitra

Symposium on Geometry Processing 2014

[Berger14] State of the Art in Surface Reconstruction from Point Clouds

M. Berger, A. Tagliasacchi, L.M. Seversky, P. Alliez, J.A. Levine, A. Sharf, C.T. Silva

EUROGRAPHICS 2014

[Averkiou15] Autocorrelation Descriptor for Efficient Co-alignment of 3D Shape Collections

Melinos Averkiou, Vladimir G. Kim, Niloy J. Mitra¹

Computer Graphics Forum 2015