

Approche sémantique de segmentation et de
recherche interactive par le contenu issu d'une
caméra de profondeur

Elliot Vanegue

1^{er} août 2016

Table des matières

| | | |
|----------|--|----------|
| 1 | Introduction | 3 |
| 1.1 | Contexte | 3 |
| 1.2 | Sujet | 3 |
| 1.3 | Problématique | 4 |
| 1.4 | Organisation | 5 |
| 2 | Etat de l'art | 6 |
| 2.1 | Segmentation | 6 |
| 2.1.1 | Scène intérieure | 6 |
| 2.1.2 | Corps humain | 6 |
| 2.2 | Reconnaissance d'objets | 7 |
| 2.2.1 | Descripteur | 7 |
| 2.2.2 | SVM | 7 |
| 2.2.3 | Bag of word | 7 |
| 2.3 | Positionnement de modèle | 7 |
| 3 | Modification des membres du corps humain | 7 |
| 3.1 | Objectif | 7 |
| 3.2 | Calcul de la distance géodésique | 7 |
| 3.3 | Calcul de descripteurs | 7 |
| 3.4 | SDK de la Kinect | 7 |
| 3.5 | Positionnement | 7 |
| 4 | Reconstruction d'un environnement intérieur | 7 |
| 4.1 | Objectif | 7 |
| 4.2 | SVM | 7 |
| 4.3 | Descripteur | 7 |
| 5 | Conclusion | 7 |

Résumé

1 Introduction

1.1 Contexte

Durant notre master IVI¹, nous avons l'occasion de réaliser un stage de fin d'étude. J'ai choisi de réaliser ce stage dans le laboratoire 3D-SAM spécialisé dans l'acquisition et le traitement d'image 3D à partir de capteur 3D de type Microsoft Kinect. Leur principaux travaux porte sur l'analyse de forme d'objet 3D et la modélisation des variation des formes dans des vidéo 3D. Réaliser mon stage dans un laboratoire était une priotité, car mes projets d'avenir ne sont pas encore tout à fait planifié et ce stage me permet de me décider sur l'environnement de travail qui me convient le mieux. Durant mes deux derniers stage en IUT et en licence 3, j'ai pu observer les atouts et les contraintes de chaque environnement, cependant le stage que j'ai effectué durant mon année en IUT relevé plus de l'ingénierie que de la recherche. Durant ce stage de deuxième année d'étude je n'ai pas eu à effectuer un état de l'art, ni à tester des méthodes proposé par d'autre laboratoire de recherche.

Ces deux années de master nous ont beaucoup appris sur le traitement d'image et la reconnaissance de forme 2D. Ces sujets mon particulièrement intéressé et je souhaite, durant ce stage, approfondir ces notions sur des types de donnée plus complet comme sur des images 3D. C'est pourquoi je réalise mon stage dans l'équipe 3D-Sam avec qui j'avais déjà travaillé sur mon projet de fin d'étude. Cette équipe dirigé par M. Daoudi comprend cinq membre permanant, un post-doctorant et quatre doctorant. Mes encadrant durant ce stage sont M. Vandeborre et M. Wannous. Ce stage ce déroule dans le cadre du projet CrABEx qui concerne la production et l'édition de produit 3D pour des applications de type loisir. L'enjeu de ce projet est d'aider les designer 3D dans la création et l'édition de ressources graphiques en suggérant des éléments approprié durant leur processus de création ou en générant automatiquement de nouvelles ressources à partir d'éléments existant.

1.2 Sujet

Comme nous pouvons le constater depuis quelques années, les approches 3D interactive sont de plus en plus présentent dans nos vies, que ce soit dans le domaine de la médecine avec l'utilisation de simulateur pour apprendre

1. Le master Image Vision Interaction est une spécialité du master informatique de l'université de Lille 1

à réaliser des opérations complexe, dans le loisir avec les jeux vidéo dont le revenu mondial en 2015 est de plus de 90 milliard de dollars ou encore dans l'industrie pour visualiser des produit. L'engouement pour cette technologie requiert des outils de plus en plus efficace et rapide permettant à des personnes de profession plutôt artistique de laisser place à leur imagination sans s'ingérer de l'aspect technique.

Le fait de pouvoir modéliser dans un monde virtuelle des objets du monde réel, et de pouvoir modifier ces objets facilement permettrait au designer de se soustraire de certaines tâches. On peut par exemple modéliser une scène ou une personne facilement grâce à une caméra Kinect, afin d'effectuer des modifications. Il serait intéressant de récupérer le modèle 3D d'une personne et de modifier quelques unes des parties de son corps avec des membre improbable comme un bras de robot. Le principe est le même dans une pièce, si nous souhaitons remplacer des meubles d'une pièce intérieur par d'autre plus improbable. L'intérêt de cela est d'éviter à un designer de devoir remodeler une personne ou une pièce existante qui serait déjà idéal pour ce que l'on souhaite réaliser.

Ce genre de technologie existe déjà pour modéliser des visages ou des gestes dans des jeux vidéo ou des films d'animation. Il y a par exemple des capture optique basé sur des caméras infrarouge avec des marqueurs réfléchissant. La Kinect fait également partie des outils de capture de mouvement, mais elle reste très peu utilisé dans le milieu professionnel.

1.3 Problématique

Les caméras 3D nous permettent d'obtenir un nuage de point de l'environnement qu'elles enregistrent. Cependant, ce type de caméra est peu précise et elles sont souvent bruité et comporte des valeurs qui n'existe pas dans le monde réel. Ce bruit dépend de plusieurs facteur comme la luminosité de l'environnement dans lequel nous effectuons l'acquisition ou tout simplement la qualité et la précisions du capteur utilisé. La première difficulté lors de ce projet est de réussir à filtrer les données de sorte à ce qu'il ne reste que les données réellement présente. L'ensemble des données ne nous intéresse pas forcément. En effet, si nous travaillons sur le corps humain nous n'avons pas besoin de l'ensemble de la pièce intérieur, ce surplus de données nous dérange lors de nos traitements. Il faut donc trouver une solution permettant de segmenter les données que nous recevons afin de ne garder que certains objets.

Une fois que les données ont été trié, nous avons besoin de reconnaître les objets présent dans notre scène. Par exemple, dans le cas du corps humain, si nous souhaitons modifier un membre comme la main il faut d'abord

savoir quel membre représente la main. La reconnaissance d'objet dépend fortement du type de donnée et nécessite généralement d'avoir une base d'apprentissage assez volumineuse.

Lors de la modification d'une partie de la scène, il est nécessaire de connaître la position de l'objet remplacé, afin de pouvoir placer le nouvel objet au même endroit et dans le même sens. Cette partie est très délicate surtout pour des membres humain, car si le repositionnement n'est pas parfait est qu'il y a un écart entre deux membres, la scène perd toute sa crédibilité et son réalisme.

Nous avons donc trois problématique majeur dans ce projet. Comment filtrer les données de manière à ne traiter que celles qui nous intéresse? Comment reconnaître un objet dans une scène 3D? Comment connaître la position exacte des objets présents dans notre scène?

1.4 Organisation

Pour répondre aux questions précédentes, je vais travailler sur deux applications mettant en avant des problématique similaire, mais avec deux façon différentes d'aborder ces problèmes. Je vais travailler dans un premier temps sur une application qui se focalise sur le corps humain. Cette application aura pour but d'identifier les membres du corps afin que l'utilisateur puisse les modifier grâce à des modèles 3D existant. Les seuls interactions de l'utilisateur sont donc de cliquer sur le membre à modifier et de sélectionner un modèle parmi ceux que l'application lui proposera. La seconde application se concentre sur une pièce intérieur comportant plusieurs objets. Ici le but est de segmenter la pièce afin de reconnaître les objets présents afin de les ajouter dans une scène. Encore une fois, l'utilisateur doit sélectionner un meuble et sélectionner le modèle 3D qui lui convient dans une liste proposé par l'application.

Pour ce projet j'utilise la caméra Microsoft Kinect v2, qui est l'une des caméras les plus utilisé dans la littérature. Le SDK fournit avec cet outil comporte une segmentation du corps humain, la position du squelette de l'utilisateur et l'algorithme Kinect fusion qui permet de construire un modèle 3D à partir du nuage de point fournit par la caméra.

Pour réaliser les deux application, je vais utiliser un ensemble de bibliothèque tel que la bibliothèque graphique de Microsoft pour la réalisation de l'interface, la bibliothèque PCL² pour les calculs sur les nuages de points et opencv pour travailler sur les images couleurs et les images de profondeur

2. Point Cloud Librairie

fournit par la Kinect.

2 Etat de l'art

2.1 Segmentation

La première étape lors de notre projet va être de segmenter les images que nous recevons de notre caméra. Les informations contenues dans une image 3D sont nombreuses et nous devons déterminer les éléments importants pour nos traitements. Dans notre scène, nous avons besoin des objets proches ou du corps de la personne en face de la caméra, mais l'environnement autour de ces objets clés n'est pas important et doit être supprimé pour gagner du temps lors de nos traitements en supprimant de l'information à traiter. Une seconde segmentation est nécessaire pour le traitement du corps humain. Pour cette étape du projet, nous devons segmenter le corps en plusieurs parties pour pouvoir par la suite les reconnaître. Si cette seconde segmentation n'est pas réalisée il ne nous sera pas possible de reconnaître les mains ou encore la tête si nous ne savons délimiter les parties du corps.

2.1.1 Scène intérieure

De nombreux travaux ont été réalisés dans la segmentation d'image 2D avant que les caméras 3D ne soient ouvertes au grand public. Les premières méthodes de segmentation reposaient sur la détection de contours comme pour la méthode de P. Arbelaez et al[1]. Leur méthode repose sur le détecteur de contours gPb qui est composé de d'un seuillage sur la luminance et sur la couleur et d'une détection de texture. La fermeture des contours se fait ensuite en utilisant les superpixels. D'autres méthodes 2D utilisent un simple seuillage en utilisant par exemple la méthode de N. Otsu[2].

2.1.2 Corps humain

La caméra Kinect avec laquelle je travaille me permet de récupérer des informations 2D, grâce à une image couleur, et 3D, grâce à une image de profondeur. Plusieurs travaux se sont

2.2 Reconnaissance d’objets

2.2.1 Descripteur

2.2.2 SVM

2.2.3 Bag of word

2.3 Positionnement de modèle

3 Modification des membres du corps humain

3.1 Objectif

3.2 Calcule de la distance géodésique

3.3 Calcule de descripteurs

3.4 SDK de la Kinect

3.5 Positionnement

4 Reconstruction d’un environnement intérieur

4.1 Objectif

4.2 SVM

4.3 Descripteur

5 Conclusion

Références

- [1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 898–916, May 2011.
- [2] “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, Jan 1979.