

République Algérienne Démocratique Et Populaire
Ministère De L'enseignement Supérieur Et De La Recherche Scientifique

UNIVERSITE MOULOU D MAMMERI DE TIZI-OUZOU



Faculté Du Génie Electrique Et D'informatique
Département D'automatique

MEMOIRE DE FIN D'ETUDES
DE MASTER ACADEMIQUE

Domaine : Sciences et Technologies

Filière : Génie Electrique

Spécialité : **Commande des systèmes**

Présenter par

HORRI Mohamed

KESSI Ali

Thème

Calcule des descripteurs dans le but de la
classification automatique
d'objet 3D

Mémoire soutenu publiquement le ...03/07/2017 devant les jurys compose de :

M Nahi Ahmed

MA/A, Président

M Chelli Takfarinas

MC/A, Encadreur

M Boudjma Fadila

MC/A, Examineur

DEDICACES

A MES CHERS PARENTS POUR LEUR SOUTIEN MORAL ET FINANCIER DURANT MES ETUDES, A MES FRERES, SŒURS, ET A LA MEMOIRE DE MES GRANDS PARENTS.

AINSI A TOUS MES FIDELES AMIS(ES), A TOUS CEUX QUE J'AIME, ET A TOUS CEUX QUI ME SONT CHERS.

KESSI ALI

DEDICACES

A MES CHERS PARENTS POUR LEURS SOUTIENS MORAL ET FINANCIER DURANT MES ETUDES, A MON FRERE, SŒURS, ET A LA MEMOIRE DE MES GRANDES PARENTS.

AINSI A TOUS MES FIDELES AMIS(ES), A TOUS CEUX QUE J'AIME, ET A TOUS CEUX QUI ME SONT CHERS.

HORRI Mohamed

Remerciements :

Nous tenons à remercier profondément Dieu, le tout puissant de nous avoir donné le courage, la volonté, ainsi que la santé pour réaliser ce modeste travail.

Nous tenons à exprimer nos vifs remerciements et nos sincères reconnaissances à **M^r CHELLI**, pour sa disponibilité, ses conseils judicieux, ses directives et ses orientations concernant notre projet de fin d'étude, tous en nous accordant sa confiance et en nous faisant profiter de sa large expérience tout au long de la réalisation de ce projet de fin d'étude.

Nous remercions les membres de jury qui ont fait l'honneur de participer au jugement de ce travail.

Nous tenons à remercier toutes les personnes qui ont contribué au bon déroulement de ce projet de fin d'étude :

- ❖ A tous les professeurs de la faculté génie électrique pour leurs conseils et leurs Orientations.
- ❖ Aux personnels du département.
- ❖ A nos parents et à toute la famille **KESSI** et **HORRI**.
- ❖ A nos frères et sœurs.
- ❖ A tous nos amis.

Tables des matières

REMERCIEMENT

TABLE DES MATIERES

LISTE DES FIGURES

ABREVIATIONS

Introduction générale	1
Chapitre I : Acquisition et représentation des images 3D	4
1. Introduction	4
2. Acquisition 3D	
2.1. Techniques passive	6
2.1.1. Triangulation passifs	6
a. Vison stéréoscopique ou stéréovision	6
b. Reconstruction à partir de mouvement (Shape form Motion)	8
c. Reconstruction à partir des contours « Shape from Silhouettes »	8
d. Shape from focus.....	10
e. Schape from Texture	10
f. Shape from shading	11
2.2. Techniques active	11
2.2.1. Triangulation actifs	11
a. Lumière structurée : point.....	13
b. Lumière structurée : ligne	14
c. Lumière structuré : motif	14
d. Triangulation laser	15
2.2.2. Temps de vol (Time of Flight : TOF)	16

3. La Kinect	17
3.1. Introduction	17
3.2. Présentation de la Kinect	18
3.3. Kinect en chiffre	19
3.4. Caractéristiques de la Kinect	19
3.5. Fonctionnement de la Kinect	20
3.5.1. Caméra couleur	20
3.5.2. Camera de profondeur	23
a. Capteur infrarouge	24
b. Source de lumière infrarouge structurée	25
c. Technique de triangulation	25
d. Précision du calcul de profondeur	27
4. Représentation 3D	28
4.1. Image de profondeur et de couleur (RGB-D)	28
4.2. Nuage de point	28
4.3. Création du nuage de point	29
4.4. Format des nuages de points	31
4.4.1. Versions PCD	31
4.4.2. Présentation du format	31
5. Conclusion	37

Chapitre II : Les descripteurs..... 38

1. Introduction	39
2. Notion de base	40
2.1. Notion de voisinage 3D	40
2.2. Estimation de la normale dans un nuage de point	40
3. Les descripteurs 3D	41
3.1. Définition d'un descripteur	41
3.2. Descripteurs locaux	43
3.2.1. Définition	43
3.2.2. Point FeatureHistograms (PFH)	44

3.2.3. FPFH :(Fast Point Feature Histograms).....	46
3.3. Descripteurs globaux	48
3.3.1. Définition	48
3.3.2. Viewpoint feature histograms (VFH)	48
3.3.3. Clustered Viewpoint Feature Histograms (CVFH)	51
4. Conclusion.....	53

Chapitre III :

1. Introduction	55
2. Définition.....	56
3. Critères pour une bonne classification.....	56
4. Les étapes de classification	56
5. Méthodes de classifications	57
5.1. Classification supervisée	57
5.2. Classification non supervisé	57
5.3. Principe de Classification non supervisé.....	57
5.4. Méthodes de Classification non supervisé	58
a. Méthodes hiérarchiques	58
b. Méthodes non hiérarchiques	58
c. Méthode des k-means	59
6. Application.....	61
6.1. Extraction des paramètres 3D.....	62
6.1.1. Acquisition.....	62
6.1.2. Traitement	63
6.1.3. Calcul de descripteur VFH	64
6.2. Classification.....	64
6.2.1. Base de données	64
6.2.2. Matériel et logiciel utilisé	65
6.2.3. Résultats	66
7. Conclusion.....	68

Conclusion générale..... 69

Référence

Annexe

Liste des figures :

Figure I. 1 : Graphe des systèmes d'acquisition.

Figure I. 2 : points de correspondance calculés sur une paire d'images stéréoscopiques

Figure I. 3 : principe de la stéréoscopie.

Figure I. 4 : Principe d'acquisition « Shape From Motion ».

Figure I. 5 : Modélisation 3D à partir de silhouettes.

Figure I. 6 : trois textures naturelles. (a) Bâtons de bois et petites feuilles (b).La surface d'un rocher.(c) Motif linéaires sur les salles.

Figure I.7 : Synthétique schap form shaliding (Zhang ,tsai, creyer et al. 1999) , (a,b) avec la lumière devant (c,d) avec la lumière avant droite (e,f) forme correspondante à partir de reconstruction d'ombrage en utilisant la technique de Tsai et Shah(1994).

Figure I.8 : principe de la triangulation active

Figure I. 9 : les trois types de la triangulation active : (a) projection point (b) ligne (c) motif

Figure I.10 : (a) Eplison (b) Nextec

figure I. 11 : Motifs de lumière structurée codés par la couleur.

Figure I.12 : Principe de la triangulation laser

Figure I. 13 : Exemple de caméra TOF

Figure I. 14 : Principe de fonctionnement des caméras Temps-de-vol

Figure I.15 : La kinect de Microsoft.

Figure I. 16 : les composants de la Kinect

Figure I. 17 : Spectre de lumière

Figure I. 18 : Capteur CMOS recouvert d'une grille photosensible de Bayer

Figure I.19 : Surface d'un capteur CMOS

Figure I. 20 : Illustration processus du système RGB

Figure I.21 : Capteur CMOS infrarouge.

Figure I.22 : Exemple de motif laser projeté par la kinect

Figure I. 23 : schéma de la déformation du motif

Figure I. 24 : Principe de la mesure de la profondeur par la kinect

Figure I.25 : image couleur à gauche et carte de profondeur colorée en fonction de la distance.

Figure I.26 : nuage de point d'un modèle 3D de voiture.

Figure 27 : Exemple d'un fichier PCD

Figure II.1 : Voisinage d'un point « p »

Figure II.2 : nuage de point (gauche), estimation de la normale (droite).

Figure II.3 : un modèle constitué d'un nuage de descripteurs locaux (gauche) et un modèle fait d'un descripteur globale (droite).

Figure II.4 : Région d'influence pour le calcul du PFH (le point rouge et ses k-voisins bleu sont entièrement interconnectés et délimité par une sphère de rayon r).

Figure II.5 : Illustration des variations angulaires calculées entre les normales de deux points lors du calcul d'un PFH.

Figure II.6 : région d'influence pour le calcul de FPFH

Figure II.7 : barycentre et vecteur normalisé

Figure II.8 : calcul des composantes angulaires pour VFH

Figure II.9 : histogramme du VFH

Figure III.1 : 1.a image couleur, 1.b : image de profondeur, 1.c : nuage de point.

Figure III.2 : Exemple d'image segmentées : (a) pomme, (b) banane

Figure III.3 : Quelques images RGB-D de la base de données RGB-D Dataset.

Figure III.4 : la projection des classes.

Abréviations :

3D	Trois dimensions ou tridimensionnels
ToF	Time of Flight
RGB	Red Green Blue
RGB-D	Red Green BLUE Depth
SDK	Software Development kit
VGA	Video Graphics Array
PCD	Point Cloud Data
PCL	Point Cloud Library
ACP	Analyse en Composantes Principales
PFH	Point Feature Histograms
FPFH	Fast Point Feature Histograms
SPFH	Simplified Point Feature Histograms
VFH	Viewpoint Feature Histograms
CVFH	Clustered Viewpoint Feature Histograms
SDC	Shape Destrribution Component
GRSD	Global Radius-based Surface Descriptor
GFPFH	Global Fast Point Feature Histograms
ESF	Ensemble of Shape Functions
RoPS	Rotational Projection Statistics
NARF	Normal Aligned Radial Feature
RIFT	Rotation Invariant Feature Transform
USC	Unique Shape Context
3DS	Shape Context
RSD	Radius-based Descriptor Surface
SI	Spin Image

Introduction générale :

La vision est sans doute notre sens le plus développé. En effet, nous nous servons tout le temps de notre système de vision pour nous déplacer, pour estimer les distances, pour identifier les personnes et les objets qui nous entourent...etc. Nous l'utilisons sans aucune difficulté, sans même y prêter attention, bien que les processus mis en jeu soient assez complexes.

La vision par ordinateur est un domaine à la frontière de l'informatique, des mathématiques, de la physique, des neurosciences, de science de l'ingénieur, de l'intelligence artificielle et de diverses autres disciplines, qui a pour but de reproduire le plus fidèlement possible la vision humaine, pour en doter les machines. Ce domaine n'a pas cessé de se développer. Il est présent dans de nombreux secteurs (industriels, médicales, robotiques, domestiques, spatiales, automobiles, vidéo-surveillance...etc.)

La vision par ordinateur 2D ne permet qu'une analyse superficielle, et pour une meilleure compréhension et interprétation d'une scène, les chercheurs ont ajouté une troisième donnée qui est la profondeur. La 3D a suscité l'intérêt de nombreux chercheurs depuis les années cinquante, mais à cette époque la 3D étaient acquise par des systèmes trop chers et assez fiables, et pour faire les traitements ils utilisent des ordinateurs impuissants, alors que les systèmes de vision par ordinateur doivent être fiables et très rapides pour pouvoir être utilisés dans des applications en temps réel.

Grâce à la naissance des machines de calcul de plus en plus sophistiquées et puissantes, et à l'avancement des techniques informatiques, certain nombre de scientifiques ont développé des nouveaux systèmes d'acquisitions 3D fiables, et avec des prix abordables, tel que la Kinect de Microsoft et la Xiton de Acer.

Dans ce projet nous allons nous intéresser à une nouvelle technologie d'acquisition nommée "Kinect" qui a été développée au début pour la console de jeux Xbox 360 afin de pouvoir contrôler les jeux par des gestes et commandes vocales, et aujourd'hui, elle est l'un des capteurs de profondeur le plus utilisé par les chercheurs en vision par ordinateur tridimensionnel et en robotique.

L'objectif de notre travail est la classification automatique d'objets dans des images tridimensionnelles avec une approche de la méthode de classification non supervisée nommée "Algorithme des k-means" et cela à l'aide de certains paramètres appelés "descripteurs 3D".

Pour cela nous avons réparti notre travail en trois chapitres, comme suit :

-Le premier chapitre : il introduit d'abord un aperçu sur les systèmes d'acquisition des images tridimensionnelles. Ensuite il présente le principe de fonctionnement et les différentes composantes de la Kinect, et en fin il présente les différentes représentations des images tridimensionnelles (carte de profondeur, nuage de points) avec le format de fichier .PCD utilisé pour les nuages de points.

-Le deuxième chapitre : il aborde quelques notions de base propres aux nuages des points tel que la notion de voisinage, l'estimation de la normale dans un nuage de points. Ensuite il définit les principaux descripteurs 3D globaux et locaux existants et leurs caractéristiques et leurs fonctionnalités.

-Le troisième chapitre : il présente, en premier lieu, une généralité sur les méthodes de classification telle que la méthode des K-means. En deuxième lieu, il présente l'application des descripteurs 3D dans le but de la classification d'objets, de même que les résultats obtenus.

Chapitre I :

*Acquisition et représentation
des images 3D*

1. Introduction :

Les systèmes d'acquisition 3D sont en pleine expansion et donnent lieu chaque année à une profusion de travaux académiques, technologiques et industriels. Ces systèmes d'acquisitions permet l'obtention d'une donnée en plus par rapport à la 2D qui est la profondeur.

Dans ce chapitre, nous allons aborder la représentation des principales techniques d'acquisition tridimensionnelle existantes, ensuite nous présenterons le principe de fonctionnement et les différentes composantes d'un nouveau capteur 3D nommé Kinect, qui est l'élément nécessaire dans notre projet, et finalement nous allons aussi présenter des différentes représentations des images 3D les plus utilisées à savoir la carte de profondeur et les nuages de points.

2. Acquisition :

Il existe une multiplication de techniques d'acquisition utilisées pour répondre aux besoins variés d'applications et aux contraintes qu'imposent le type de chaque scène [1].

Plusieurs organismes de ce type peuvent être trouvés dans la littérature, bien que notre étude se porte uniquement sur la mesure sans contact plus précisément sur la réflexion optique, à savoir les méthodes passives et actives.

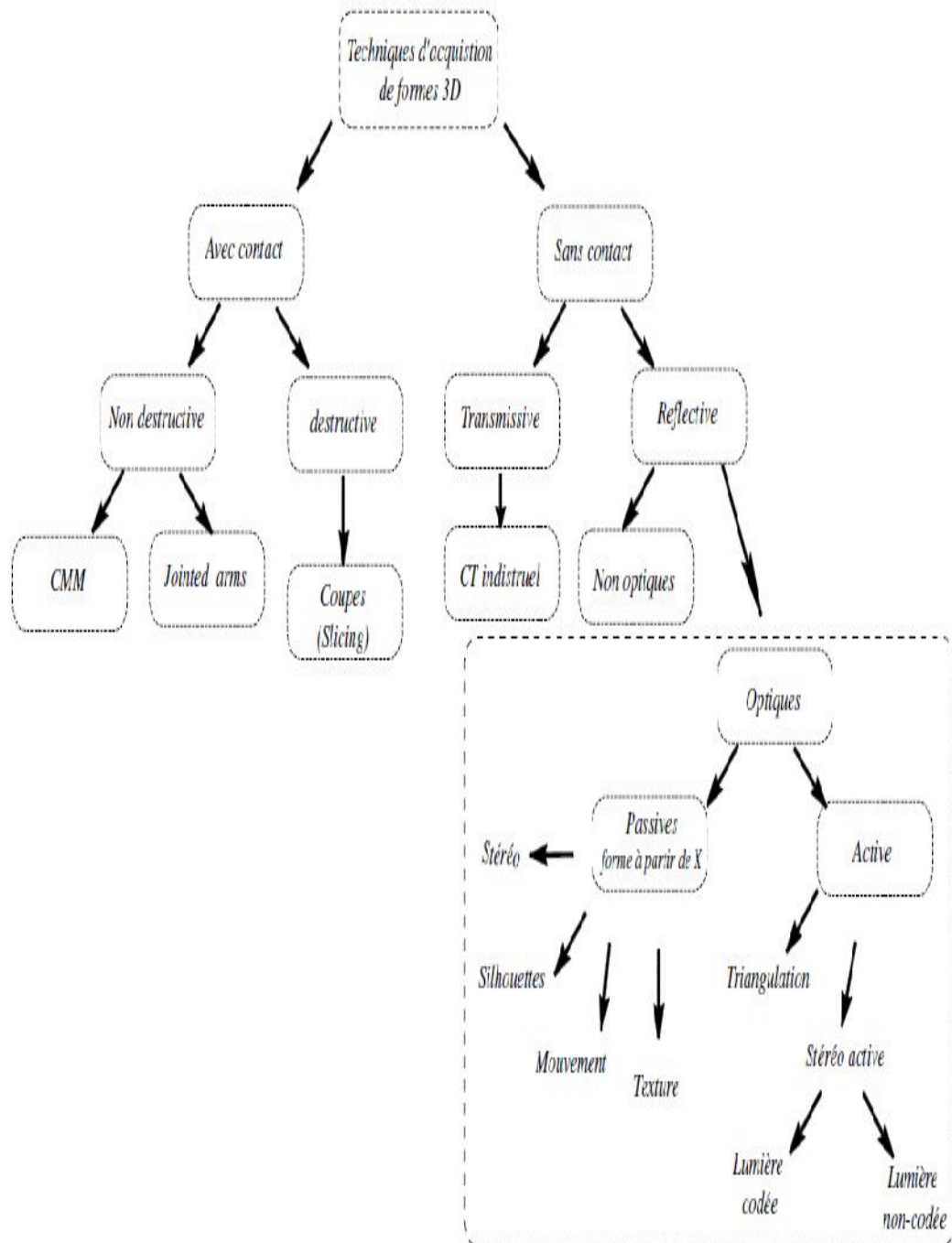


Figure I.1 : Graphe des systèmes d'acquisition[1].

2.1. Techniques passive :

2.1.1. Triangulation passives :

La triangulation passive utilise uniquement un ou plusieurs systèmes d'acquisition d'images. Elle est basée sur la réflexion du flux lumineux ambiant (lumière naturelle), et ne nécessite aucun rayon lumineux émis par la caméra. Cette technique exploite une autre caractéristique pour extraire l'information 3D : mouvement, vues multiples....

a. Vision stéréoscopique ou Stéréovision :

La vision stéréoscopique est la méthode passive la plus largement étudiée. Le principe tend à se rapprocher du fonctionnement de la vision humaine : deux caméras font l'acquisition d'une même scène à partir de deux points de vue différents et la répétition de ce processus pour plusieurs points donne la forme 3D. Une mesure n'est réalisable qu'après calibrage de caméras dans un repère unique, c'est-à-dire qu'il faut connaître les paramètres intrinsèques (focale et taille des pixels) et les paramètres extrinsèques (position et orientation) de chacune des caméras. Après l'acquisition de la paire d'image, l'étape primordiale pour la reconstruction tridimensionnelle est l'appariement ou mise en correspondance. La figure I.2 illustre l'appariement de certains points caractéristiques (angles, coins, contours) sur deux images prises par un système stéréoscopique [2].

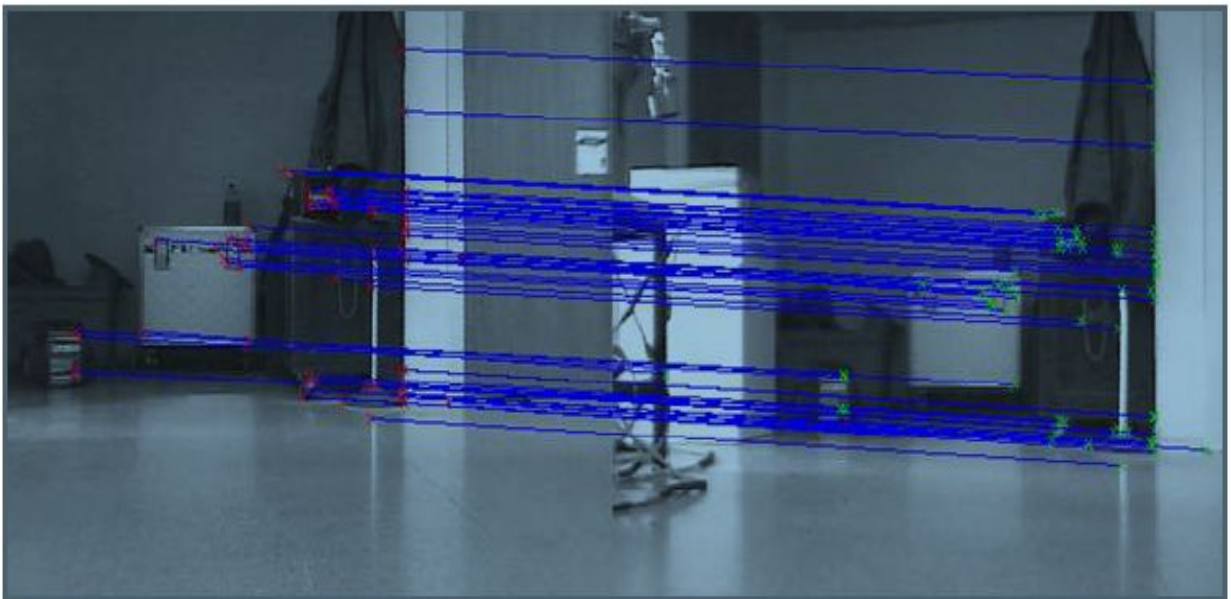


Figure I.2: points de correspondance calculés sur une paire d'images stéréoscopiques [2].

- Expliquons son fonctionnement :

Si deux images de la même scène sont acquises avec des angles différents, alors la profondeur de la scène crée une disparité géométrique entre elles. Notons ε cette fonction de disparité. Alors, ε et la fonction de profondeur z sont liés par la relation suivante :

$$z = \frac{\varepsilon}{b/h} \quad (\text{I.1})$$

Ou b/h est un coefficient de stéréoscopie dépendant seulement des conditions d'acquisition. b est la distance entre les centres des caméras et h est la distance entre la scène et le système de caméras.

La précision d_z de la mesure de profondeur est donc liée à la précision $d\varepsilon$ de la mesure de disparité :

$$d_z = \frac{dz}{d\varepsilon} = \frac{z}{\varepsilon} \quad (\text{I.2})$$

Tout ceci est représenté sur la figure I.3

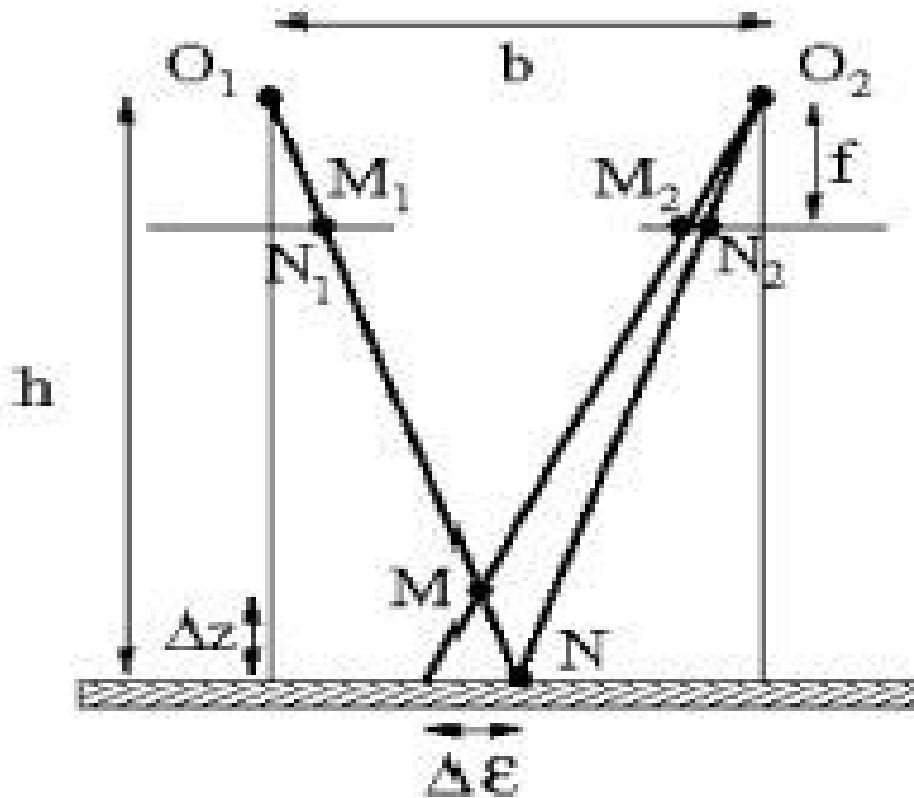


Figure I.3 : principe de la stéréoscopie [3].

Pour une précision donnée, plus le coefficient b/h est grand plus l'erreur de profondeur d_z est petite. Mais un grand coefficient b/h signifie aussi qu'il y a beaucoup de changement, entre les deux images de la paire stéréoscopique, ce qui implique plus de difficultés dans le processus d'appariement. Il faut donc faire un compromis entre une erreur de profondeur petite et une erreur de disparité petite, c'est-à-dire un processus d'appariement fiable [3].

b. Forme à partir du mouvement (Shape from Motion) :

Cette méthode permet de trouver l'information de surface à partir du mouvement relatif estimé entre l'objet et la caméra. L'acquisition est réalisée par une caméra en mouvement réalisant une prise d'images successives, soit par un objet en mouvement par rapport au système d'acquisition. Globalement, la reconstruction est réalisée en calculant le déplacement entre les caméras par la recherche des points caractéristiques [4].

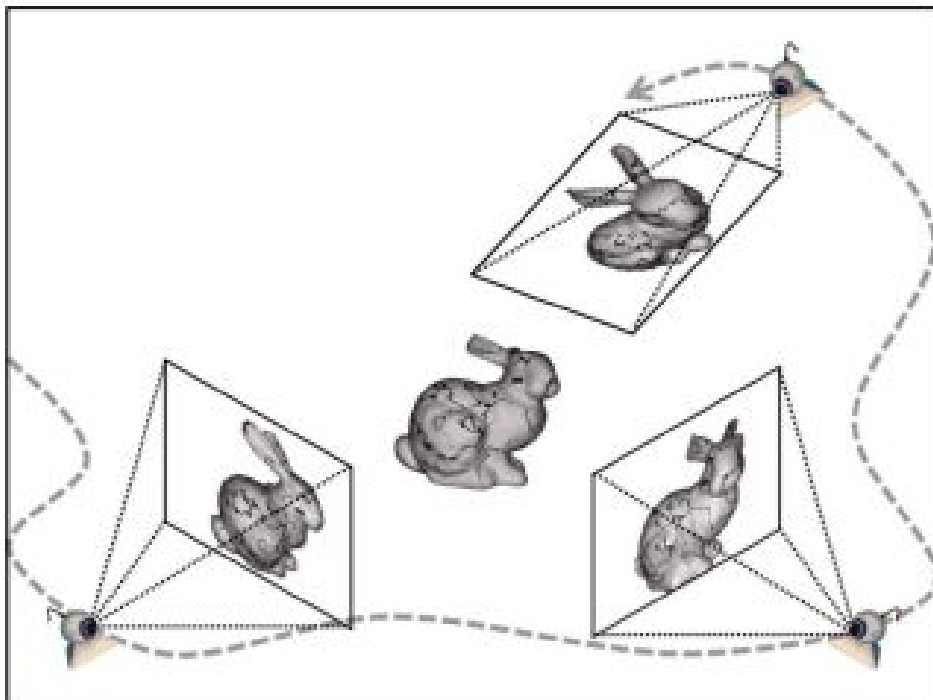


Figure I.4 : Principe d'acquisition « Shape From Motion ».

c. Reconstruction à partir des contours « Shape from Silhouettes » :

Cette méthode utilise comme donnée de départ des images prises au même instant par plusieurs caméras statiques d'une même scène. Sur chaque image, la silhouette de l'objet

étudié est extraite. Ensuite, une reconstruction 3D de l'objet étudié est calculée à partir de ces silhouettes.

Nous pouvons regrouper les méthodes de reconstruction à partir de silhouette en deux catégories : les approches volumétriques et les approches surfaciques.

Dans une approche volumétrique, l'espace est discrétisé en une grille de voxels. La forme 3D peut être vue comme l'ensemble des voxels qui sont projetés à l'intérieure des silhouettes sur les différentes caméras. Cette approche est consommatrice en calcul à cause de la discrétisation de l'espace. De plus, avec les premières approches, la décision de dire si un voxel appartient ou non à la surface 3D est délicate en pratique car les silhouettes ne sont pas exactes.

Dans une approche surfacique, la forme 3D est reconstruire comme un polyèdre. Ce polyèdre est obtenu par l'intersection des projections 3D des silhouettes dans l'espace. Cette étape de reconstruction de polyèdre est réputée complexe mais aboutit à des résultats plus précis que les méthodes volumétriques.

Cette technique à plusieurs limites : afin d'obtenir un modèle 3D de bonne qualité, un grand nombre de photographiques est nécessaire. Et si l'objet présente des surfaces concaves, elles n'apparaîtront pas sur la silouhaites et ne seront pas reconstruites.



Figure I.5 : Modélisation 3D à partir de silhouettes.

d. Shape from focus :

Cette technique s'apparente également au terme de « Depth from Focus ». Le principe de base est la mesure du flou sur une image. En utilisant des systèmes optiques à profondeur de champ très faible, il est alors possible de déduire la profondeur des points qui apparaissent nets à l'image en connaissant la distance focale. La reconstruction 3D d'une pièce se fait ensuite par sectionnement optique, c'est-à-dire en faisant l'acquisition de plusieurs images [2].

e. Shape from Texture :

L'objectif de cette technique est de mesurer l'orientation de la surface à partir de l'exploitation des informations de la texture des objets. Les variations de cette texture sur l'image permettent d'estimer la forme de la surface observée à partir d'une seule image. Le champ de la normale ainsi obtenu permet d'interpréter la forme de l'objet. La précision de la mesure reste faible [5].

Une texture régulière étant caractérisée par un motif de base se répétant périodiquement.

Cette technique est donc valable seulement si la surface de l'objet ne présente pas de discontinuités.

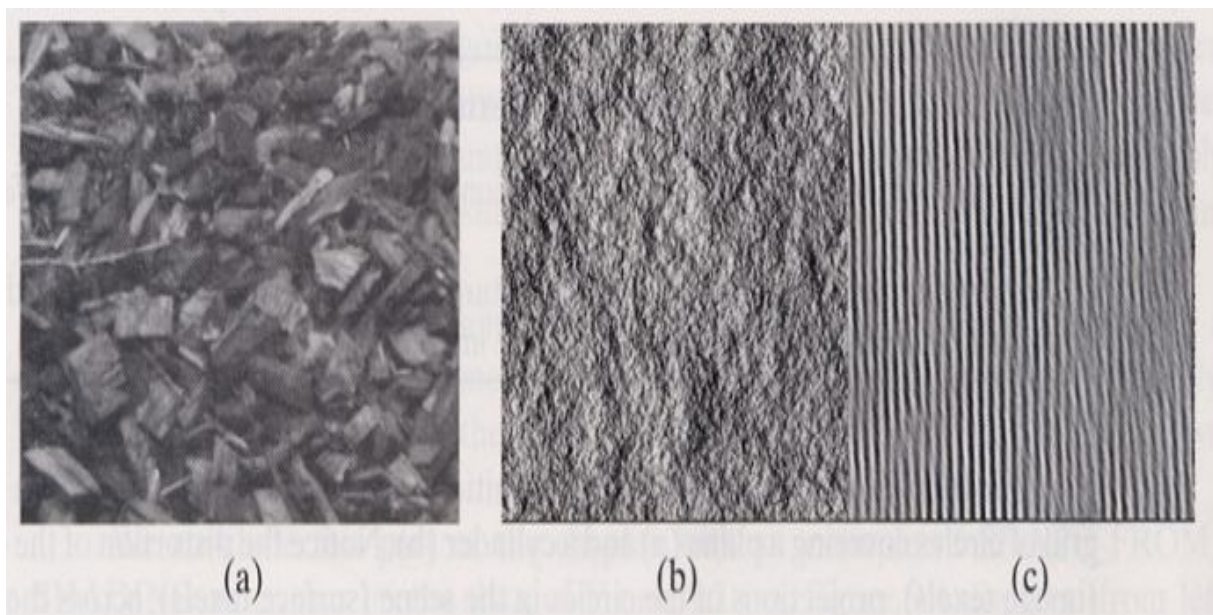


Figure I.6: trois textures naturelles. (a) Bâtons de bois et petites feuilles (b) La surface d'un rocher. (c) Motif linéaires sur les sables.

f. Shape from shading :

Horn est le premier à passer la problématique dans les années 1970. Une revue plus détaillée des méthodes a été donnée, par la suite, par Zhang et al [6]. L'enjeu de la technique est d'extraire la forme 3D d'un objet à partir de la réflectance de la surface supposée connue et de la carte d'intensité donnée par l'image acquise. En effet, le niveau d'intensité donnée par chaque pixel dépend de plusieurs facteurs qui sont : l'éclairage de la scène, la forme de la surface, sa réductance et la position du capteur.

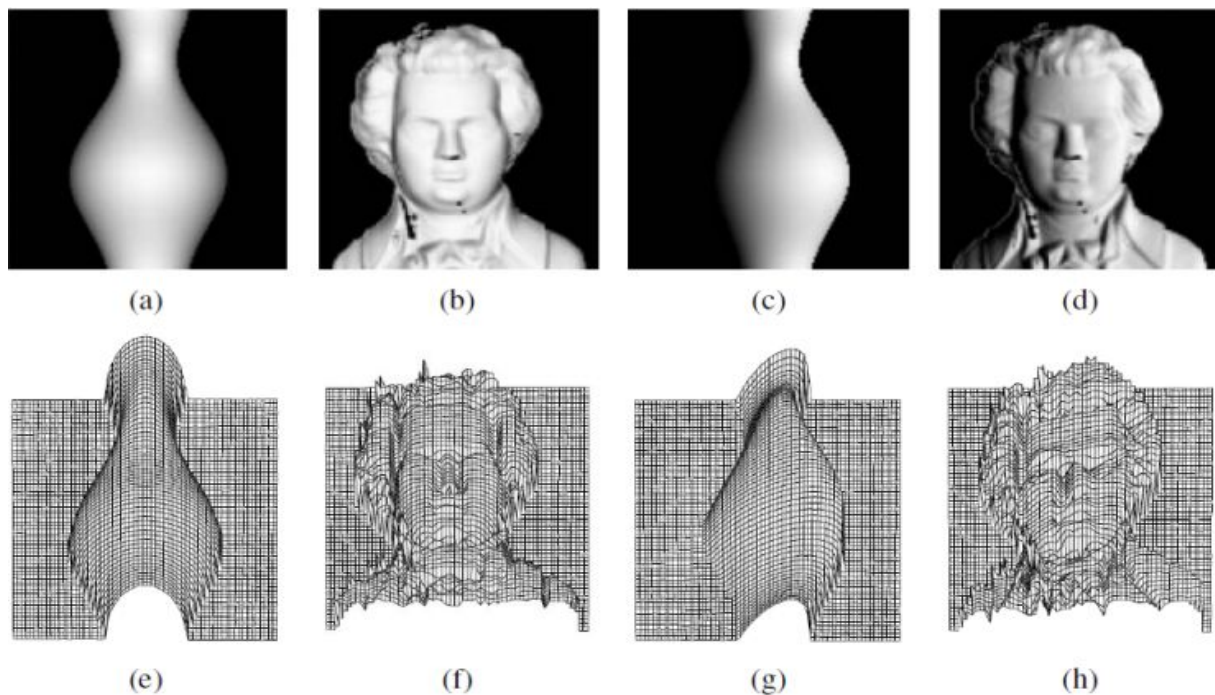


Figure I.7: Synthétique schapformshalding (Zhang ,tsai, creyer et al. 1999) , (a,b) avec la lumière devant (c,d) avec la lumière avant droite (e,f) forme correspondante à partir de reconstruction d'ombrage en utilisant la technique de Tsai et Shah(1994).

2.2. Techniques active :

2.2.1.Triangulation actifs :

La technique de triangulation active se distingue de la technique passive sauf qu'une des deux caméras de cette dernière est remplacée par un éclairage structuré qui participe à la reconstruction tridimensionnelle de l'objet.

Le principe de cette technique est de projeter sur l'objet des motifs connus pour mesurer la profondeur par une simple trigonométrie. Cette technique est appelée triangulation car le point laser, l'appareil photo et l'émetteur laser forment un triangle [7].

Dans l'exemple donné dans la figure(I.8), O c'est le centre optique du capteur, α est le centre de l'image, f est la focale. Les orientations du projecteur θ et du capteur φ , ainsi que la distance D entre le projecteur et le capteur sont connues.

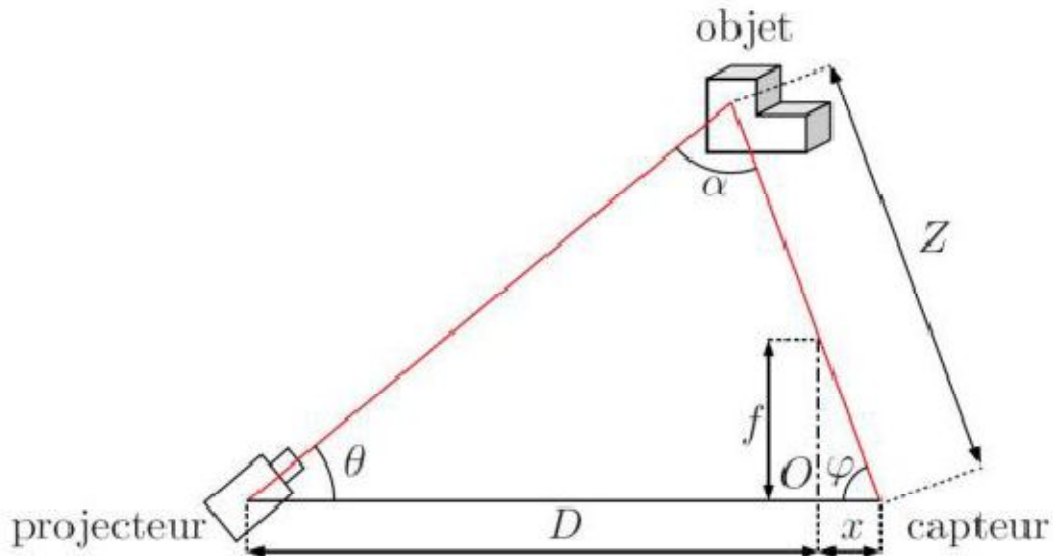


Figure I.8: principe de la triangulation active

Pour déterminer la profondeur Z en fonction de la déviation x lue sur l'image, « la loi des sinus » :

$$\frac{D + X}{\sin \alpha} = \frac{D + X}{\sin(\theta + \varphi)} = \frac{Z}{\sin \theta}$$

$$\text{Avec : } \varphi = \arctan \frac{f}{x} \quad (\text{I.3})$$

$$Z = (D + x) \frac{\sin \theta}{\sin(\theta + \varphi)}$$

Les diverses méthodes de triangulation active peuvent être classées suivant le type du motif lumineux projeté : point, ligne, motif ou laser.

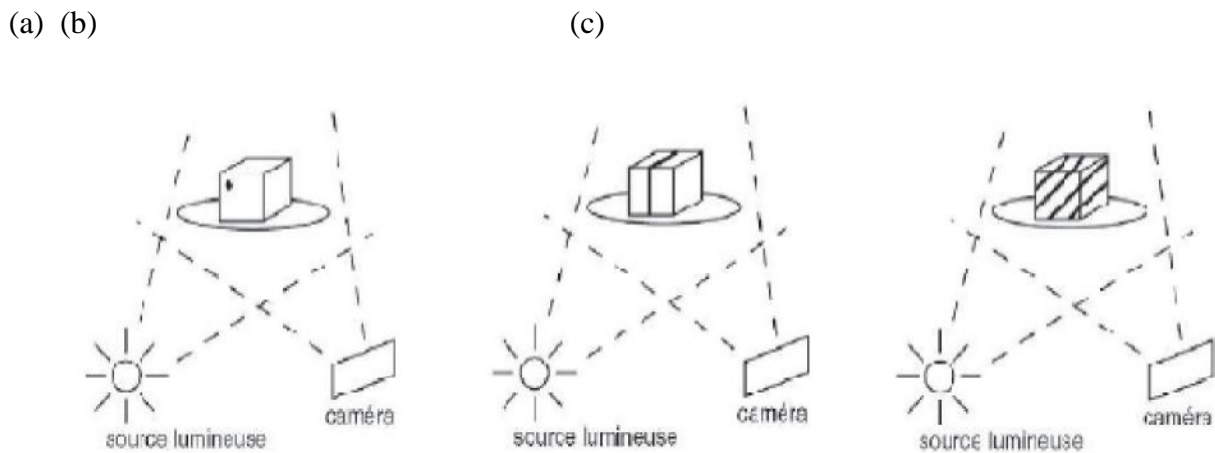


Figure 1.9 : les trois types de la triangulation active : (a) projection point (b) ligne (c) motif

a. Lumière structurée : point

Cette technologie nécessite une source laser point, la figure I.10 montre deux exemples des têtes laser point. Le projecteur lumineux et le capteur doivent être mobiles, afin de couvrir entièrement l'objet [8].

Un point laser est projeté sur l'objet à numériser et acquis par la caméra. Le système étant préalablement calibré, l'orientation et l'inclinaison spatiale de la caméra et du laser sont connues. Ceci permet, par triangulation, de déterminer la distance entre la caméra et l'objet aux niveaux du point projeté. Le principal inconvénient de cette méthode est le nombre très élevé des points à acquérir pour obtenir une reconstruction acceptable de l'objet. Ce qui signifie que le capteur doit être mobile [4].

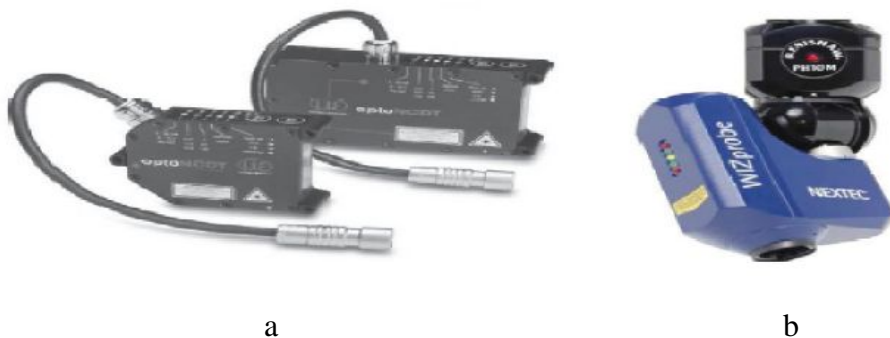


Figure I.10. : (a) Eplison (b) Nextec [10].

b. Lumière structurée : ligne

Les scanners qui utilisent cette technologie sont plus rapides que ceux qui utilisent la projection point. La longueur de la ligne et le nombre de point sur la ligne varient en fonction du modèle du capteur (entre 100 et 500 point par ligne) [8].

Dans ce cas, les relations de triangulation ne sont plus utilisées pour déterminer la profondeur mais une étude de la déformation de la ligne par l'objet permet de remonter à cette information. Un balayage de l'objet est nécessaire pour la numérisation complète, d'où le nom « scanner 3D » couramment utilisé pour dénommer cette méthode. On retrouve ce type de système sur des bras poly articulés ou des appareils de scanning utilisent des miroirs rotatifs pour balayer la scène. L'utilisation principale concerne le domaine industriel pour de la rétro-conception ou du contrôle qualité et, de plus en plus, le domaine du patrimoine pour la numérisation d'œuvres d'art. Un autre problème survient lorsque la ligne n'est pas entièrement visible par la caméra (phénomène appelé occlusion), ce qui entraîne un problème de mise en correspondance [4].

c. Lumière structurée : motif

Ce qui distingue cette technologie, par rapport aux autres citées auparavant, est qu'elle utilise la projection de points, de lignes, de grilles ou de motifs codés. A la différence de la triangulation laser ligne, le motif projeté recouvre entièrement une face de l'objet, ce qui permet l'acquisition immédiate d'une surface. Il en résulte une plus grande rapidité d'acquisition. Afin d'appliquer la triangulation sur chacun des motifs, la première étape consiste en l'identification de chacune des raies projetées. Cette étape peut entraîner des difficultés lorsque les motifs projetés présentent des discontinuités [9].

Dans le cas de la projection d'un motif couleur, la texture de l'objet peut être acquise en étudiant la différence de teintes entre la couleur émise et la couleur observée.

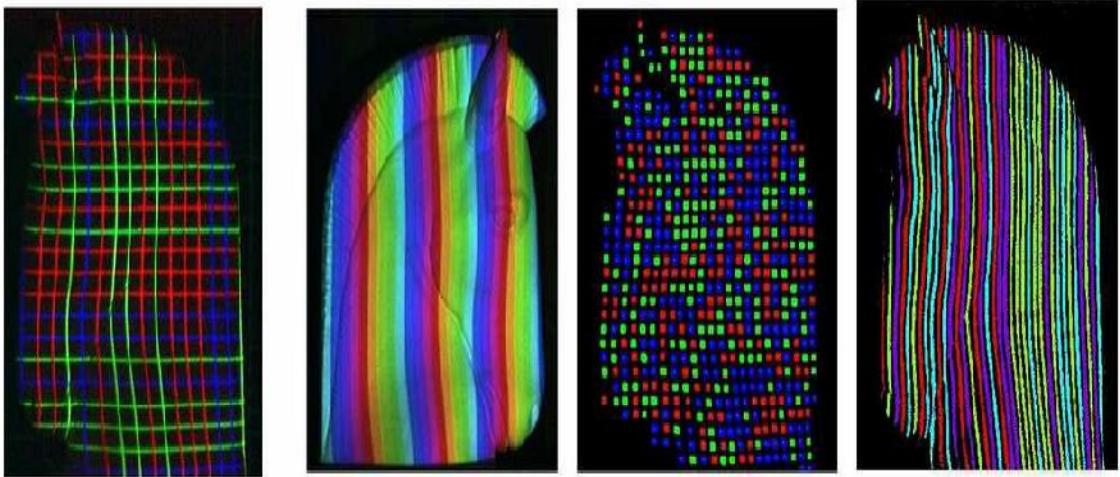


Figure I.11 :Motifs de lumière structurée codés par la couleur.

d. Triangulation laser :

Les scanners 3D par triangulation laser utilisent deux dispositifs : un émetteur laser et un appareil photographique. Ce genre de scanners est appelé « scanners par triangulation » car l'émetteur laser, l'appareil photo et le point laser projeté forment un triangle.

L'émetteur laser forme un angle avec l'axe de l'appareil photo et l'angle ainsi que la distance entre l'axe de l'appareil photo et le point laser projeté peuvent être calculés.

Généralement l'émetteur laser émet une ligne qui balaye l'objet à scanner afin d'accélérer la numérisation.

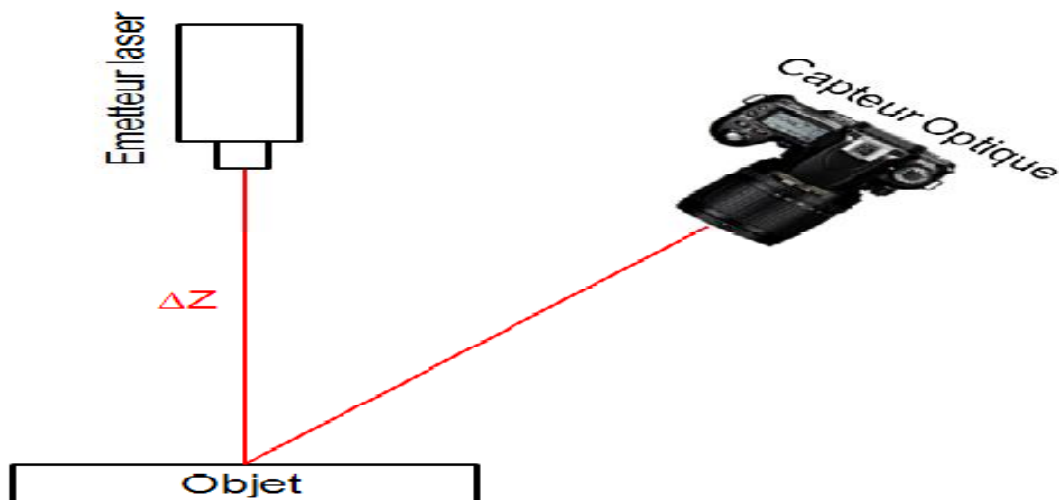


Figure I.12 :Principe de la triangulation laser

2.2.2. Temps de vol :

Le scanner par temps de vol (Time-Of-Flight, TOF) est un système actif de mesure point par point, basé sur le principe de la télémétrie (précision de l'ordre du millimètre).

a. Composant d'une caméra à temps de vol :

Un appareil à temps de vol est constitué des composants suivants :

- Source lumineuse : il illumine la scène.
- Capteur d'image : c'est le cœur de la caméra TOF. Chaque pixel mesure le temps que la lumière a pris pour se rendre de l'unité d'éclairage à l'objet.

b. Principe de fonctionnement des caméras Temps-de-vol :

Un émetteur envoie une onde de nature lumineuse sur la surface à mesurer. L'onde est réfléchi sur un récepteur qui mesure son intensité ainsi que son temps d'aller-retour. Du fait de la haute sensibilité des éléments électroniques qu'elle contient elle mesure des durées de l'ordre de la picoseconde [2].

Cette technologie est utilisée lorsque l'objet est très grand (bâtiment), ces scanners ont une portée pouvant aller jusqu'à 300 mètres et ont une précision de l'ordre du millimètre [7].



Figure I.13 : Exemple de caméra TOF

Les coordonnées d'un point 3D (X, Y, Z) dépendent de trois données : l'angle horizontal, l'angle vertical et la distance entre le scanner et l'objet. Les valeurs des angles sont

données par l'orientation de la source laser. La distance d entre le capteur et un point est calculée par la relation suivante :

$$d = v \frac{\Delta t}{2} \quad (\text{I.4})$$

Où v est la vitesse du signal, et Δt est le temps mis par le signal pour effectuer l'aller-retour, et donc $\Delta t/2$ est le temps mis par l'impulsion laser pour atteindre l'objet.

Ces systèmes peuvent également être utilisés dans l'industrie, pour le contrôle de pièces de grandes dimensions [8].

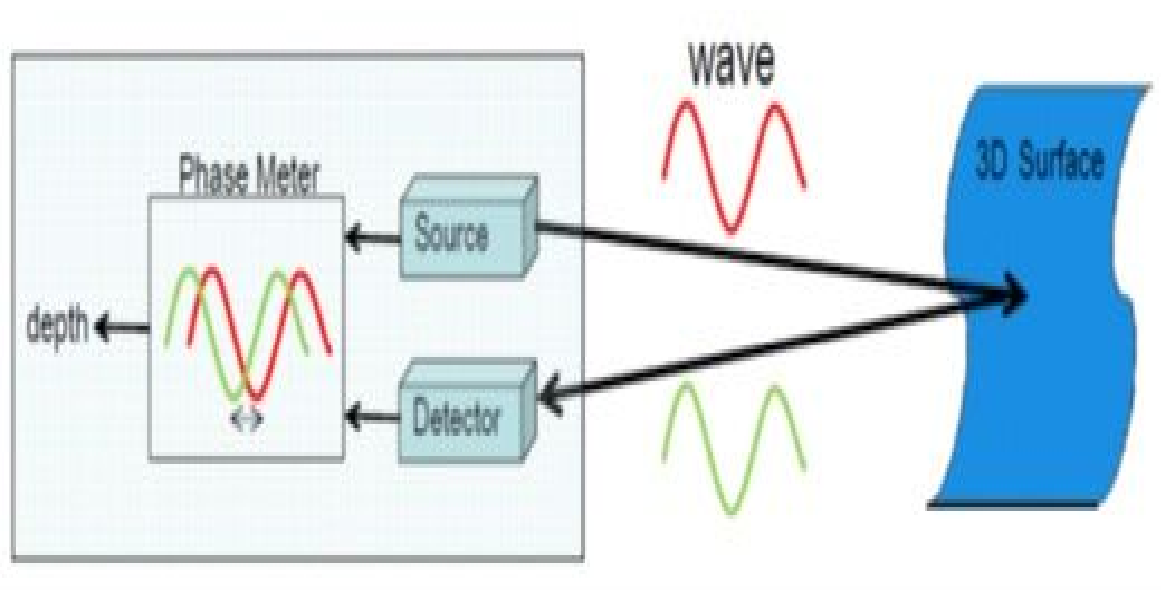


Figure I.14 : Principe de fonctionnement des caméras Temps-de-vol

3. La Kinect :

3.1. généralités:

La Kinect est une camera qui utilise des techniques d'interaction développée par la société israélienne « PrimeSense » pour la console du salon Xbox 360 de Microsoft en 2010. Elle a longtemps été nommée par son nom de code "Projet Natal" avant d'être officialisée et baptisée juste avant un grand salon dédié aux jeux vidéo, l'E3, sous le nom Kinect. Le mot Kinect" est issu des mots anglais "Kinetic" (que nous pouvons traduire par "cinétique") et "connect" (que nous pouvons traduire aussi par "connecter").

La Kinect avait remporté un franc succès lors de sa commercialisation individuelle. Elle a trouvé rapidement d'autres applications que le secteur de jeux vidéo. Elle a pu être amenée à se développer dans de nombreux secteurs comme l'industrie robotique, système de sécurité, médecine....

Le faible coût de la Kinect et son poids peu encombrant en font l'un des capteurs 3D les plus utilisés pour le développement de nouvelles applications toujours plus interactives.

3.2. Présentation de la Kinect :

La Kinect de Microsoft, initialement développée pour l'industrie des jeux vidéo (console de jeux Xbox), elle permet aux utilisateurs de contrôler les jeux par des gestes et commande vocale. Elle est ensuite utilisée dans d'autres domaines comme la vision par ordinateur tridimensionnel et en robotique. Cela est principalement dû à son faible coût qui en fait l'un des capteurs 3D les plus abordables (comparé aux paires stéréo et aux caméras à temps de vol).

Ce périphérique de Microsoft se forme principalement d'une barre horizontale constituant l'élément principale de la technologie Kinect et contenant une caméra RGB fournissant des images couleur et un capteur de profondeur pour détecter la profondeur de chaque pixel de l'image comme illustre la figure I.16. À partir de ces deux images, la reconstruction 3D de la scène est rendue possible.



Figure I.15 : La Kinect de Microsoft.

3.3. Kinect en chiffre :

La Kinect est sortie le 4 novembre 2010 en Amérique du Nord et le 10 novembre 2010 en Europe. Cette nouvelle technologie ne coûte que 150€ Plus de 8 millions d'exemplaires de la Kinect ont été vendus partout dans le Monde, 2 mois après sa sortie, dont un million en seulement 10 jours. Le 11 mars 2011, la Kinect entre au livre Guinness des records comme étant « l'accessoire high-tech le plus vendu dans un court laps de temps » avec 10 millions d'unités vendues, soit une moyenne officielle de 133 333 Kinects vendues chaque jour à travers le monde.

Aujourd'hui, plus de 100 millions d'exemplaires ont été vendus.

3.4. Caractéristiques de la Kinect :

•Capteurs :

-Caméra couleur et caméra de profondeur

-Micro à reconnaissance vocale

•Motorisation :

-Ensemble motorisé pour suivre les déplacements

Inclinaison : +/-27 degrés

•Champ de vision :

-Champ de vision horizontal : 57 degrés

-Champ de vision vertical : 43 degrés

-portée du capteur : 0,8m-4,0m

•Flux de données :

-320 * 240, 16 bits à 30 images par seconde pour le capteur infrarouge

- 640 * 480, 32 bits à 30 images par seconde pour le capteur couleur

-Audio 16 bits à 16 kHz pour les 4 microphones

•**Système de reconnaissance physique :**

-Jusqu'à 6 personnes et 2 joueurs actifs (4 joueurs actifs avec le SDK 1.0)

-20 articulations par squelette

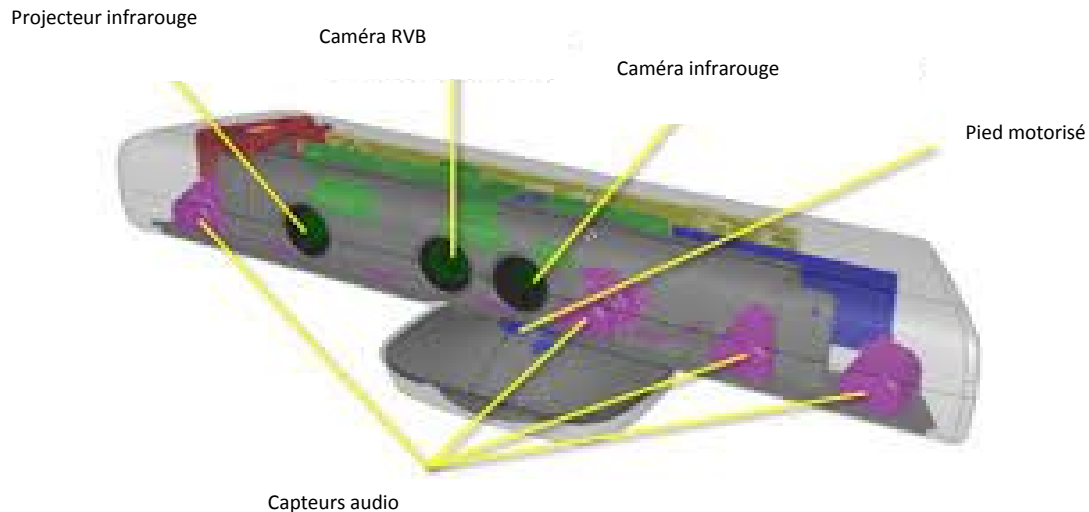


Figure I. 16 : les composants de la Kinect

3.5. Fonctionnement de la kinect :

3.5.1. Caméra couleur :

La caméra couleur est la première des deux capteurs embarqués dans la technologie kinect. Elle se situe au centre de la barre horizontale (voir figure I.16). Elle dispose d'un capteur photographique CMOS, qui permet de capturer les images à une vitesse de 30 images par seconde avec une résolution VGA (Vidéo Graphiques Array) de 640*480 pixels. Le système RGB est simple : il combine les 3 couleurs primaires telles que le rouge, le vert et le bleu. Suivant l'association de ces 3 couleurs, nous pouvons obtenir une image blanche ou un panel infini de couleurs qui se situe dans le spectre visible de l'œil humain [10]. La figure (I.17) illustre le principe du système RGB dans le processus d'acquisition d'une image couleur.

Longueur d'onde :

-Rouge : 630 nm

-Vert : 560 nm

-Bleu : 470 nm

IL:

-Rouge : 0,35 μ A

-Vert : 0,05 μ A

-Bleu : 0,30 μ A

Vf : 0,5 v

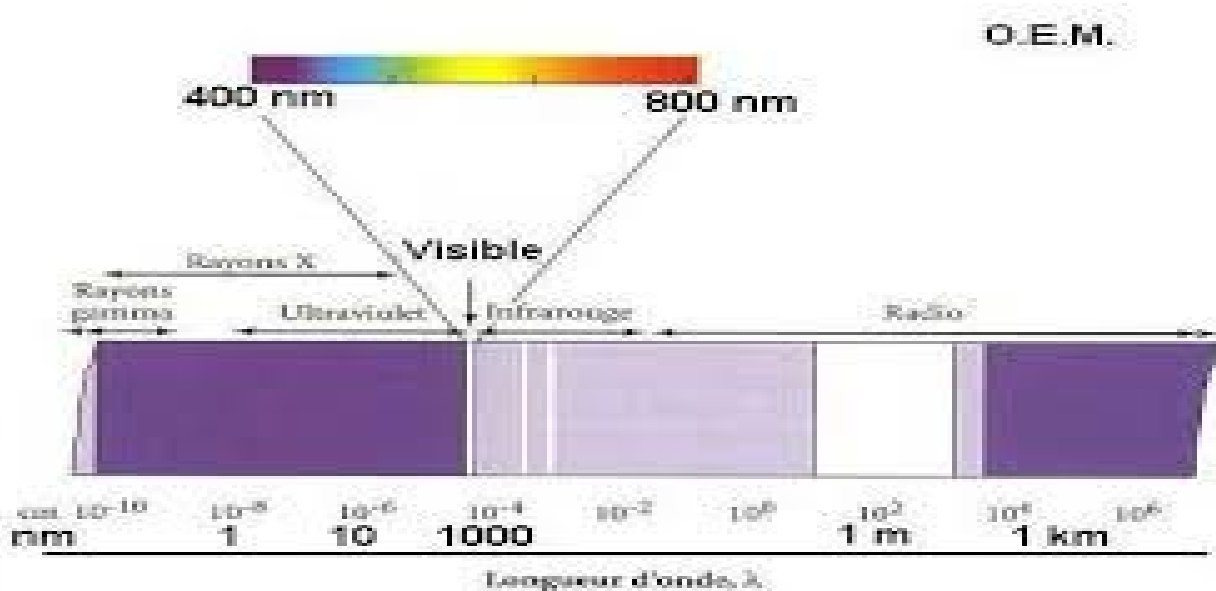


Figure I. 17 : Spectre de lumière

Le capteur CMOS est un composant électronique sensible à la lumière qui converti un rayonnement infrarouge en un signal analogique. Ce signal est ensuite numérisé par un convertisseur analogique-numérique.

Ces capteurs sont le résultat de l'intégration de cellules composées d'une photodiode et d'une logique d'amplification puis d'obtention. Ils sont produits par des techniques classiques de micro-électronique et de ce fait, ils peuvent avoir des résolutions très importantes. Comme beaucoup de capteur, les capteurs CMOS pour l'imagerie couleur sont

associés à un filtre coloré et un réseau de lentilles dû à la faible surface de la photodiode, seule zone sensible.

Le capteur CMOS se compose d'une matrice de cellules photosensibles qui conçoivent leur charge en les transférant elles-mêmes au convertisseur, là ou d'autres capteurs comme CCD se compose d'une matrice de cellules photosensibles qui transfèrent la charge vers un collecteur qui transfère à son tour l'ensemble des charges vers le convertisseur. Le capteur CMOS a l'avantage de ne pas transmettre une simple charge, Son photo site intègre un amplificateur de tension lui permettant de convertir lui-même la charge électrique reçue [10].

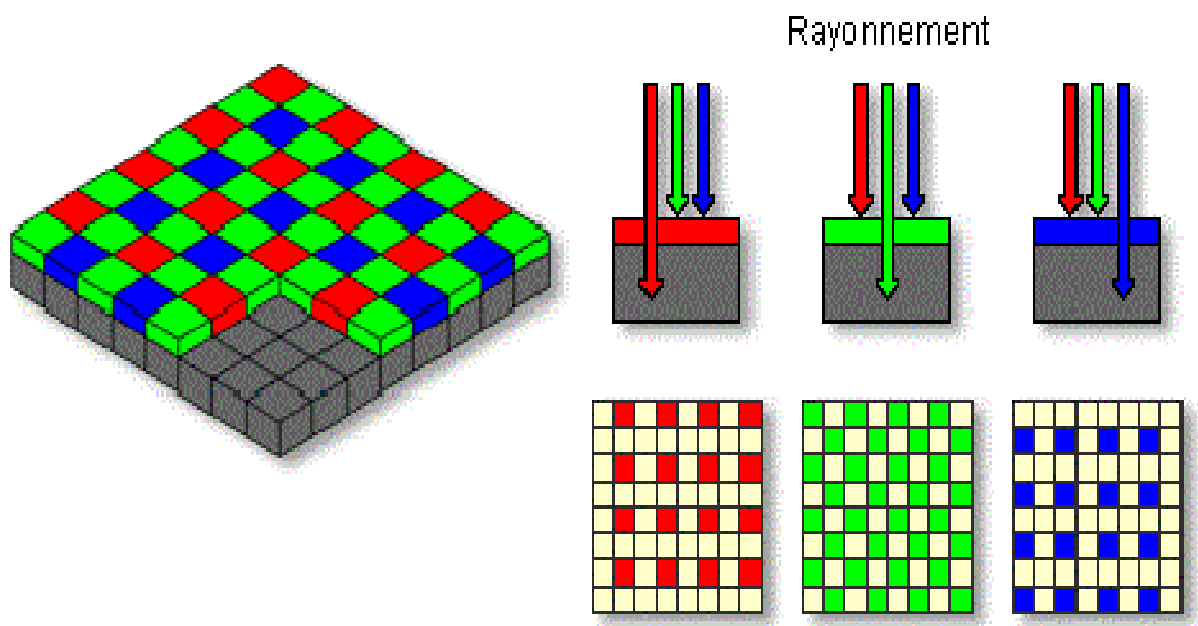


Figure I. 18 : Capteur CMOS recouvert d'une grille photosensible de Bayer

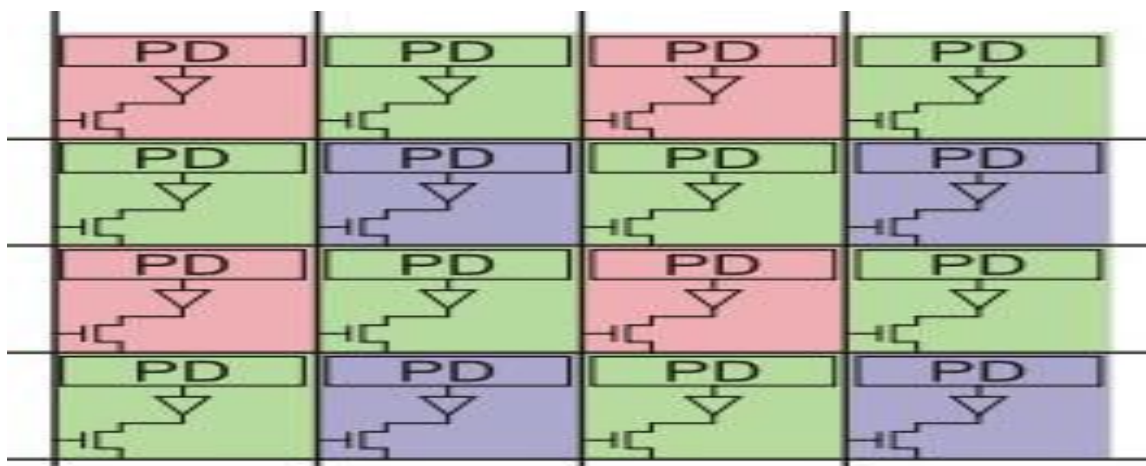


Figure I.19 : Surface d'un capteur CMOS.

Les capteurs CMOS fonctionnent par champ de photodiode (PD) disposées en grille de Bayer chacune étant sensible à une seule couleur primaire (rouge, vert, bleu) (figure I.19). C'est un champ de «0 »ou « 1 » qui recouvre alors le capteur. C'est donc une combinaison des valeurs de chaque photodiode qui permet d'obtenir la couleur estimée de chaque pixel.

La figure suivante résume le principe du système RGB et les processus pour l'acquisition d'une image avec un haut contraste et de très bonnes couleurs [10].

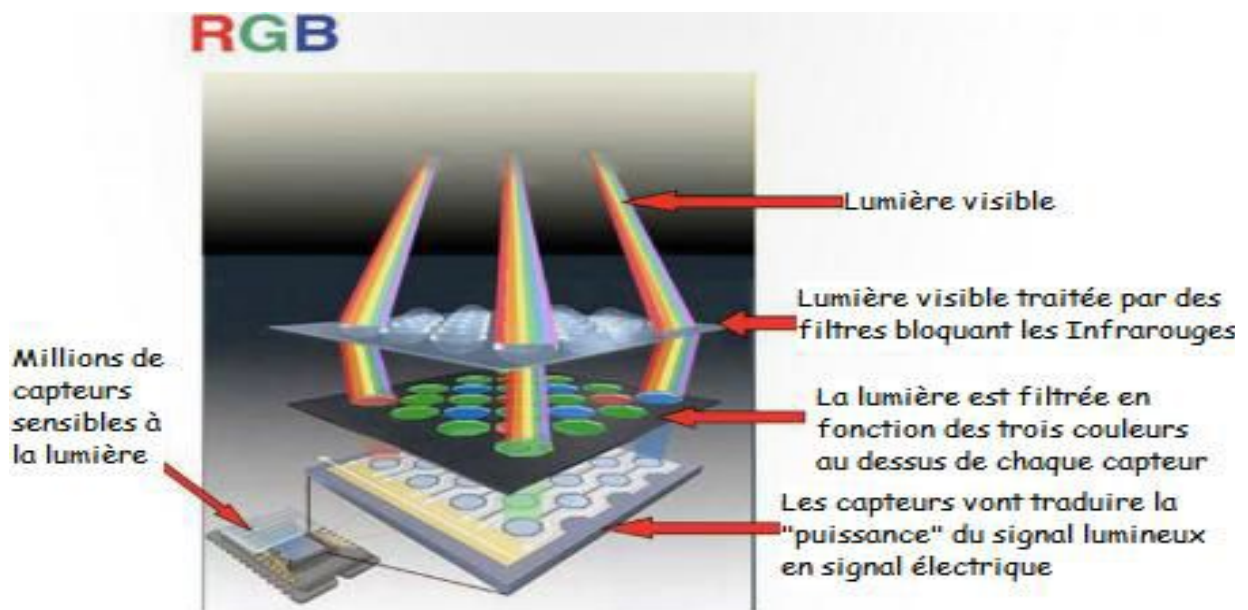


Figure I. 20 : Illustration processus du système RGB

3.5.2. Camera de profondeur :

La caméra de profondeur repose sur la projection d'un motif de lumière structurée infrarouge sur la scène observée et une caméra infrarouge qui capture une image du motif projeté.

a. Capteur infrarouge :

La caméra de profondeur utilise un capteur CMOS infrarouge qui fonctionne quasiment de la même manière qu'un capteur CMOS RVB sauf qu'il laisse uniquement

passer les infrarouges grâce à un filtre. Ce capteur est conçu suivant une architecture dite hybride constituée de deux zones (voir (figure I.21) comme suit :

- Une zone de détection constituée de $M \times N$ détecteurs élémentaires dont le rôle de chaque détecteur consiste à transformer le flux de photons infrarouge incident en un signal électrique.
- Une zone de traitement du signal obtenue à l'aide d'un circuit intégré réalisé en silicium, ce circuit appelé circuit de lecture sert à collecter et traiter le signal de chaque détecteur et à le convertir dans un format exploitable par l'électronique de mesure (en générale une tension).

La technique utilisée ici, pour connecter point à point la mosaïque de détecteurs élémentaires au circuit de lecture en silicium, est un contact électronique par l'intermédiaire de microbilles d'indium [10].

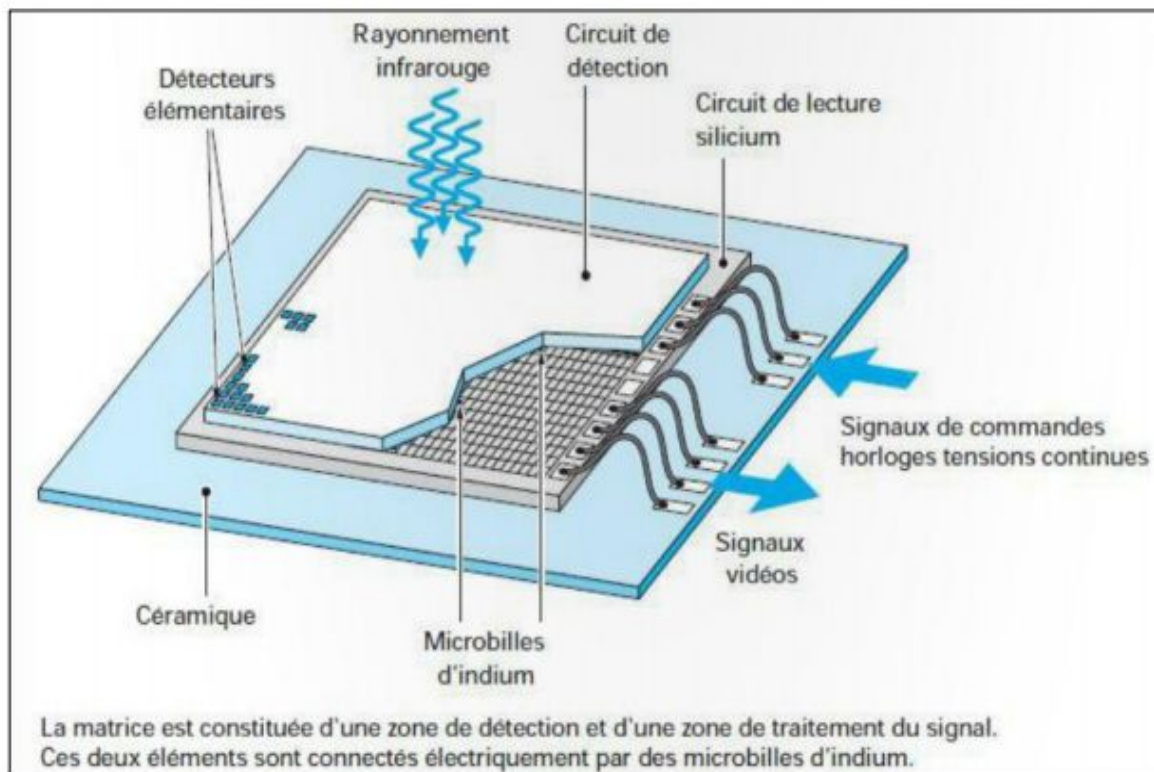


Figure I.21 : Capteur CMOS infrarouge.

b.Source de lumière infrarouge structurée :

La source de lumière infrarouge émet un faisceau unique qui est subdivisé en faisceaux multiples par réseau de diffraction afin de créer un modèle constant de mouchetures

projetées sur la scène (fig. I.22). Cette tendance est capturée par la caméra infrarouge et comparée avec un modèle de référence.



Figure I.22 : Exemple de motif laser projeté par la Kinect.

c. Technique de triangulation :

Le principe de lumière structurée permet d'obtenir l'information de profondeur d'objets présents dans une scène en projetant un motif lumineux sur les objets et en analysant la déformation de ce motif au contact des objets.

La Kinect projette un motif infrarouge, invisible à l'œil nu dans la scène. Au contact des objets de la scène, le motif se déforme comme illustre la **figure I.23**.

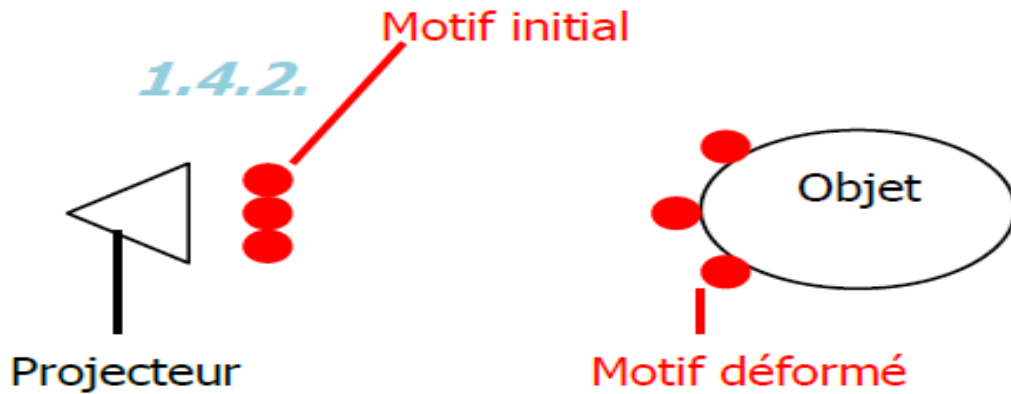


Figure I. 23 : schéma de la déformation du motif

Le capteur infrarouge, étant légèrement décalé du projecteur infrarouge, peut ainsi capturer la déformation du motif pour le comparer avec un motif de référence. Le motif de référence est obtenu en capturant un plan à une distance connue par la kinect, puis capturé et stocké dans la mémoire de la kinect. Cette configuration de la kinect est réalisée en usine avant la commercialisation. Quand le point du motif projeté sur un objet dont la distance au capteur est plus petit ou plus grande que celui du plan de référence, alors le point du motif dans l'image infrarouge décale dans la direction de la ligne de base entre le laser projecteur et le centre focal de la caméra infrarouge. Ce déplacement observé par la caméra est appelé la disparité [11].

Il est possible d'estimer la profondeur de chaque point du motif par un calcul de disparité entre le motif et son image :

$$z = \frac{BXf}{d} \quad (I.5)$$

Avec d la disparité, B la Baseline (la distance entre le centre de la source et le centre du capteur), f la focal de la caméra et z la profondeur.

La figure I.24 illustre le principe de la mesure de la profondeur de la kinect.

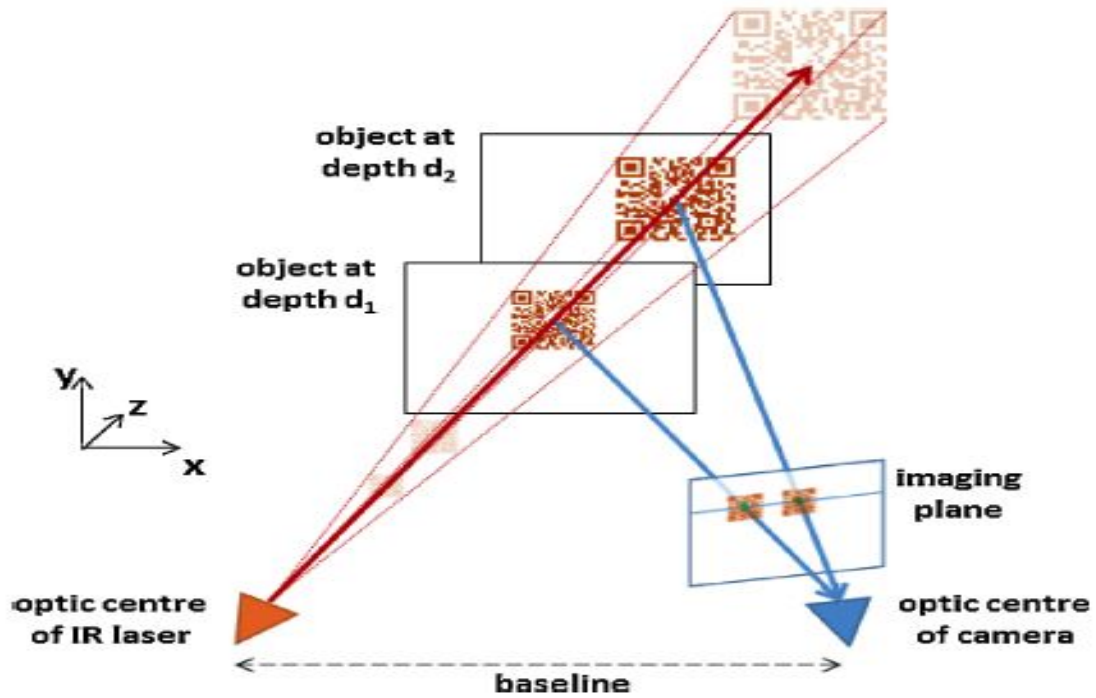


Figure I. 24 : Principe de la mesure de la profondeur par la Kinect [11].

Plusieurs auteurs ont analysé les performances de la Kinect pour la mesure de la profondeur. Il est, a priori, d'utiliser la Kinect dans un environnement entièrement sombre. Cependant, la caméra 3D n'est pas utilisable le jour, en extérieur, du fait de la présence des longueurs d'onde du proche infrarouge dans la lumière émise par le soleil, ces rayons perturbent le calcul des profondeurs. Ces différents travaux montrent que la Kinect est très adaptée pour la cartographie intérieure lorsque les distances de mesure sont faibles et inférieures à 3.5 mètre [12]. A cette distance, la Kinect offre de meilleures performances qu'une caméra à temps de vol. D'autre part, Khoshelham et al. Dans [12] montrent que l'erreur sur la mesure de la profondeur croît de manière quadratique avec la distance de mesure.

d. Précision du calcul de profondeur :

Le motif infrarouge est projeté sur les objets de la scène. Plus les points du motif déformé sont éloignés les uns des autres, plus la précision du calcul de profondeur est faible. Ainsi, la précision des informations dépend de la distance de l'objet par rapport à la caméra.

4. Représentation 3D :

4.1. Image de profondeur et de couleur (RGB-D) :

Une image de profondeur ou carte de profondeur, émanant de la même lentille que l'image d'intensité, puisque c'est le traitement des faisceaux infrarouges qui permet de réaliser une mesure de distance entre le capteur et l'objet. La carte de profondeur, qui se présente également sous la forme d'une image en niveaux de gris codée sur 16 bits, est une matrice ayant pour taille la résolution de la lentille infrarouge. Pour plus de visibilité, la carte de profondeur peut donc être colorée.

Plus le pixel a une valeur élevée, plus le point est proche de la camera de profondeur et inversement, plus le pixel a une valeur faible plus il est loin. Un pixel noir (valeur 0) signifie qu'aucune mesure n'a pu être faite. Les coordonnées x et y du pixel donnent sa position en abscisse et en ordonnée [13].

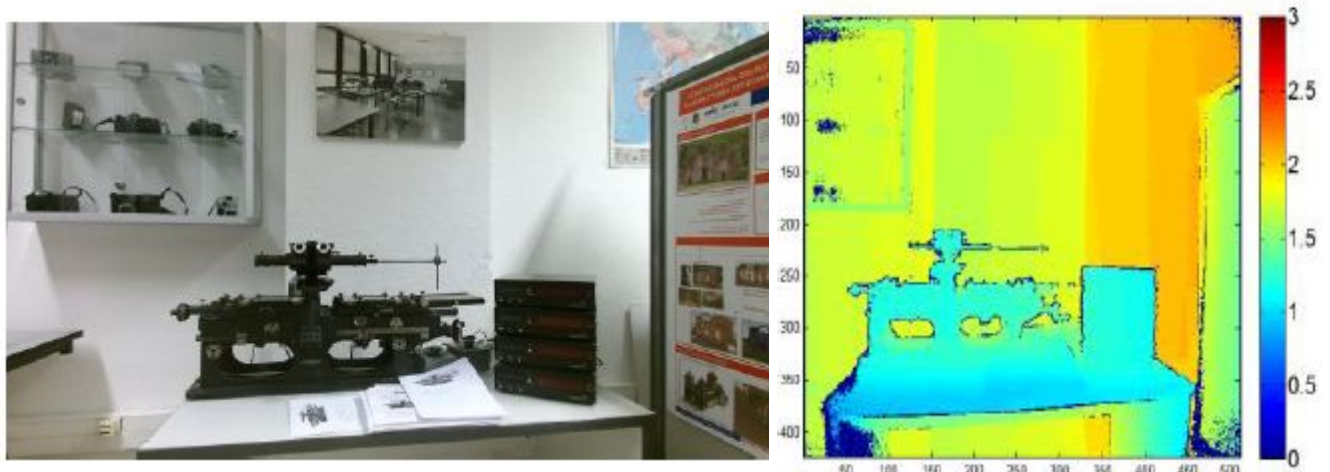


Figure I.25 : image couleur à gauche et carte de profondeur colorée en fonction de la distance.

4.2. Nuage de point :

Le nuage de point représente une collection de mesures 3D permettant de décrire la surface d'une scène. Ces points sont considérés comme des échantillons de la surface de l'objet plus ou moins précis de l'environnement et sont arrangés tel que chaque point possède une coordonnée en x, y et z , avec z représentant la profondeur. Selon le type de capteur, le nuage de point peut être organisé (structuré) ou désorganisé (non structuré). le nuage de point

est couramment soumis à corriger ou compléter ces données .Certains capteurs (kinect par exemple) permettent d'obtenir de l'information supplémentaire sur les points mesurés, comme par exemple, la valeur d'intensité de retour du laser ou la couleur de la surface. Dans ce cas, le nuage de points peut être en quatre démentions [14].

De manière générale, un nuage de points 3D peut être représenté par :

$$P = [p_1, p_2, p_3 \dots \dots p_n] \quad (\text{I.6})$$

Ou un point p_i est décrit par ses coordonnées :

$$p_i = [x_i, y_i, z_i]^T \quad (\text{I.7})$$

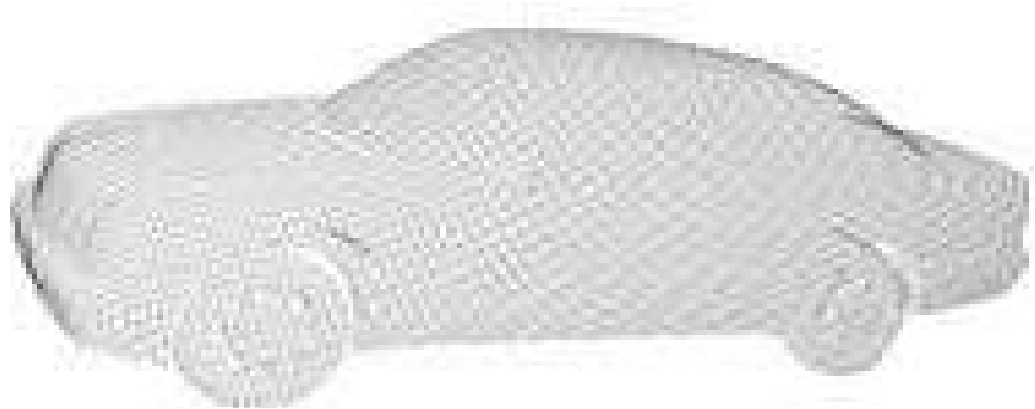


Figure I.26 : nuage de point d'un modèle 3D de voiture.

4.3. Création du nuage de point :

Avec les informations acquises par la kinect nous pouvons créer pour chaque vue le nuage de point correspondant comme suit :

-Pour chaque vue de l'acquisition d'une image couleur 2D, la kinect fournit une carte de profondeur de taille 640 x 480.

-La carte de profondeur donne pour chaque pixel de l'image couleur sa valeur de profondeur correspondante

En utilisant ces informations et les paramètres intrinsèques de la caméra trouvée, nous pouvons créer un nuage de point 3D correspondant à la vue.

Les paramètres intrinsèques sont la distance focal en abscisse F_x , la distance focale en ordonnée F_y , le coefficient de distraction en abscisse C_x , et le coefficient de distorsion en ordonnée C_y .

On peut ainsi transformer chaque pixel 2D de coordonnées (i, j) en point 3D de coordonnées (x, y, z) grâce aux formules suivantes :

$$Z = \text{valeurdeprofendeurdupixel}(i, j)$$

$$X = (i - C_x) * Z / F_x \quad (1.8)$$

$$Y = (j - C_y) * Z / F_y$$

Nous réalisons ce calcul pour chaque pixel de la carte de profondeur pour obtenir un nuage de point 3D correspondant. En effectuant la même opération pour toutes les vues, nous obtenons les nuages de points 3D pour chaque vue.

Remarque :

Un nuage de points organisé est le nom donné aux nuages de points qui ressemblent à une image organisée (ou matrice), où les données sont réparties en lignes et en colonnes. Les avantages d'un ensemble de données organisée est qu'en connaissant la relation entre les points adjacents (par exemple pixels), les opérations de voisins les plus proches sont beaucoup plus efficaces, ce qui accélère le calcul et abaisse le coût de certains algorithmes dans PCL.

Le format PCD offre les avantages suivants :

- La possibilité d'enregistrer et de manipuler de nuage de points organisés. Ce procédé est très important pour les applications en temps réel.
- Les types de données binaires sont chargés et sont enregistrés plus rapidement dans le disque dur.
- La capacité d'enregistrer des types de données différents (tous les types primitifs : char, short, int, float, double), donne une flexibilité aux nuages de points.

Un avantage supplémentaire est qu'en contrôlant le format de fichier, nous pouvons mieux l'adapter à PCL, et obtenir ainsi la meilleure performance par rapport aux applications.

4.4. Format des nuages de points :

Il existe plusieurs formats de fichier pour décrire les données 3D de nuage de point, et parmi ces formats le format de fichier PCD (Point Cloud Data) qui est un nouveau format créé par les développeurs spécialement pour les nuages de point. Il vient compléter les formats de fichiers existants qui, pour une raison ou une autre, ne supportent pas certaines des extensions que PCL apporte au traitement de nuage de point.

PCD n'est pas le premier type de fichier créé pour soutenir les données 3D. Parmi les formats qui ont été créés pour décrire le nuage de point acquis en utilisant des scanners laser nous avons : PLY, STL, OBJ et X3D.

4.4.1. Versions PCD :

Les formats de fichiers PCD peuvent avoir différents numéros de révision, avant la publication du point Cloud Library (PCL) version 1.0. ceux-ci sont numérotés avec PCD_V_X (par exemple, PCD_V5, PCD_V6, PCD_V7, etc.) et représentent les numéros de version 0.x pour le fichier PCD. Le point d'entrée officiel pour le format de fichier PCD dans PCL devait cependant être la version 0.7 (PCD_V7).

4.4.2. Présentation du format :

Entête :

Chaque fichier PCD contient un entête qui identifie et déclare certaines propriétés des données de nuages de points stockés dans le fichier. L'entête d'un PCD doit être codé en ASCII.

Depuis la version 0.7, l'entête PCD contient les entrées suivantes :

- ❖ **VERSION** : spécifie la version du fichier PCD
- ❖ **FIELDS** : spécifie le nom de chaque dimension/champ qu'un point peut avoir.

Exemple :

```

FIELDS x y z                # XYZ data
FIELDS x y z rgb# XYZ + colors
FIELDS x y z normal_xnormal_ynormal_z# XYZ + surface normals
FIELDS j1 j2 j3            # moment invariants
...

```

- ❖ **Size** : Spécifie la taille de chaque dimension en octets.

Exemple:

- ✓ Unsigned char / ombre a 1 octet
- ✓ Unsigned short / court a 2 octets
- ✓ Unsigned int / int / float dispose de 4 octets
- ✓ Doubles dispose de 8 octets

- ❖ **TYPE** : spécifie le type de chaque dimension comme un caractère. Les types acceptés sont :

- ✓ **I** : représente types int8 (signé char), Int 16(courte),et int32 (int).
- ✓ **U** : représente les types non signé unité8 (unsignedchar), uint16 (unsigned short), uint32(unsignedint).
- ✓ **F** : représente les types de floatteurs.

- ❖ **COUNT** :

Spécifie le nombre d'éléments que possède chaque dimension. Par exemple, x données à généralement 1 élément, mais un descripteur de fonctionnalité comme le VFH à 308, si COUNT n'est pas présent, le nombre de toutes les dimensions est fixé à 1.

- ❖ **WIDTH** :

Spécifie la largeur de l'ensemble de données de nuages de points dans le nombre de points.

WIDTH a deux significations :

- ✓ Il peut spécifier le nombre total de point dans le nuage pour les ensembles de données non organisés.
- ✓ Il peut spécifier la largeur (nombre total de point dans une rangée) d'un nuage de points organisé.

❖ **HEIGHT** : spécifie la hauteur de l'ensemble de données des nuages de points dans le nombre de points. HEIGHT a deux significations :

- ✓ Il peut spécifier la hauteur (nombre total de lignes) d'un ensemble de données de nuages de points organisés ;
- ✓ Il est mis à 1 pour les ensembles de données non organisés (ainsi utilisés pour vérifier si un ensemble de données est organisé ou non).

Exemple :

WIDTH640 # *Image-like organized structure, with 640 rows and 480 columns,*
HEIGHT480 # *thus 640*480=307200 points total in the dataset*

Exemple :

WIDTH307200
HEIGHT1 # *unorganized point cloud dataset with 307200 points*

❖ **VIEWPOINT** :

Spécifie un point de vue de l'acquisition pour les points dans l'ensemble de données. cela pourrait être plus tard utilisé pour la construction entre les différents systèmes de coordonnées, ou pour aider avec des fonctionnalités telles que les normales de surface, qui ont besoin d'une orientation cohérente.

Les informations de point de vue sont spécifiées comme traduction (txtytz) + quaternion (qwqxqyqz). La valeur par défaut est :

VIEWPOINT0001000

❖ **POINTS :**

Spécifie le nombre total de points dans le nuage. Depuis la version 0.7 ,son but est un peu redondant, donc nous nous attendons à ce qu'il soit retiré dans les futures versions.

Exemple :

POINTS307200# *le nombre total de points dans le nuage*

❖ **DATA :**

Spécifie le type de données que les nuages de points stockent. Dans la version 0.7, deux types de données sont pris en charge : Ascii et binaire.

A noté :

Les entrées d'entête doivent être spécifiées précisément dans l'ordre ci-dessous, à savoir :

VERSION

FIELDS

SIZE

TYPE

COUNT

WIDTH

HEIGHT

VIEWPOINT

POINTS

DATA

Remarque :

Les fichiers PCD offre une meilleure flexibilité et une vitesse plus grande que tous les autres formats.

Exemple d'un fichier PCD :

```
# .PCD v.7 - Point Cloud Data file format
VERSION .7
FIELDS x y z rgb
SIZE 4 4 4 4
TYPE F F F F
COUNT 1 1 1 1
WIDTH 213
HEIGHT 1
VIEWPOINT 0 0 0 1 0 0 0
POINTS 213
DATA ascii
0.93773 0.33763 0 4.2108e+06
0.90805 0.35641 0 4.2108e+06
0.81915 0.32 0 4.2108e+06
0.97192 0.278 0 4.2108e+06
0.944 0.29474 0 4.2108e+06
0.98111 0.24247 0 4.2108e+06
0.93655 0.26143 0 4.2108e+06
0.91631 0.27442 0 4.2108e+06
0.81921 0.29315 0 4.2108e+06
0.90701 0.24109 0 4.2108e+06
0.83239 0.23398 0 4.2108e+06
0.99185 0.2116 0 4.2108e+06
```

Figure 27 : Exemple d'un fichier PCD

5. Conclusion :

Nous avons présenté dans ce chapitre, un aperçu des méthodes d'acquisition tridimensionnelle de l'espace à partir d'une ou plusieurs images. Il existe, deux méthodes qui sont qualifiées pour résoudre ce problème : la première dite la méthode active et la deuxième dite la méthode passive.

La première méthode nécessite pas beaucoup de calcul, mais en revanche elle est trop coûteuse financièrement (matériels sophistiqués). Par contre, la deuxième méthode nommée passive ne nécessite que des caméras pour l'acquisition des images.

L'objectif commun des méthodes actives et des méthodes passives est la reconstruction 3D de la scène observée, c'est-à-dire le calcul des coordonnées 3D de chaque point de l'ensemble des points constituant la scène dans l'espace.

Nous avons présenté aussi une caméra 3D très récente « la Kinect » qui présente une évolution ingénieuse dans le domaine des systèmes d'acquisition 3D. Elle est destinée aux grands publics, contrairement aux autres systèmes d'acquisition qui ont été utilisés sauf par les professionnels du domaine, et cela grâce à son prix particulier abordable, et sa manipulation facile.

La Kinect est utilisée dans de nombreux projets et différents domaines d'application comme la robotique, la reconnaissance de forme et de mouvement, la modélisation 3D ..., et cela grâce aux données qu'elle fournit, à savoir les différentes représentations 3D (la carte de profondeur et le nuage de point) que nous avons présentées à la fin de ce chapitre.

Chapitre II :

Les descripteurs 3D

1. INTRODUCTION :

Les images 3D comportent de nombreuses informations géométriques il est donc nécessaire d'étudier les structures de données qui servent à caractériser un ensemble de points multidimensionnels. Les nuages de point sont les représentations des formes 3D les plus utilisées, car les plus simples et les plus commodes, stockent des coordonnées X, Y, Z et la couleur RGB. Mais aucune de ces propriétés n'est assez unique. Dans deux balayages séquentiels, deux points pourraient partager les mêmes coordonnées malgré leur appartenance à différentes surfaces. Cependant, la structure de nuage de point n'a pas la régularité des images 2D ou 3D, ce qui complique l'analyse des formes.

Lors de l'analyse de formes rigides représentées par des nuages de points, il faut considérer des propriétés résistantes aux translations, aux rotations et aux changements d'échelle. Pour des objets non rigides, il faut considérer des propriétés résistantes aux changements de posture, voire aussi aux changements topologiques. Par exemple, il faut reconnaître une personne qu'elle soit assise, debout, couchée, avec ou sans les mains dans les poches. Cette contrainte est essentielle pour garantir la performance de la reconnaissance automatique des formes. Plusieurs recherches ont été consacrées aux descripteurs qu'ils sont utilisés pour analyser, représenter, et ensuite reconnaître et distinguer les formes.

La reconnaissance d'objet repose généralement, sur des descripteurs qui représentent l'objet de manière unique et facilitent son identification dans la scène. Parmi les nombreux descripteurs de surface proposés dans la littérature, certaines sont des extensions de descripteurs 2D, tandis que d'autres ont été spécifiquement développées pour caractériser des nuages de points 3D. Ces descripteurs doivent par conséquent être robustes aux certaines transformations telles que la rotation, la translation, et le bruit sur les surfaces...etc.

Dans ce chapitre nous allons d'abord voir quelques notions de bases propre aux nuages de points tel que la notion de voisinage, l'estimation de la normale dans un nuage de points, Ensuite nous allons définir les principaux descripteurs 3D globaux et locaux existants, leurs caractéristiques et leurs fonctionnalités.

2. Notions de bases :

2.1. Le voisinage 3D :

Dans le contexte de descripteurs, le concept de voisin est un concept à la fois primitif et fondamental. Nous dirons alors que l'ensemble de ces points forme le voisinage de point p si, nous identifions tous les points situés à une distance inférieure à un rayon « r » de ce point p qui représente l'origine.

Le choix du rayon « r » est selon l'application voulu parce que, si nous choisissons un petit « r », nous prenons le risque d'avoir des estimations très erronées si les données sont contaminées par le bruit. Cependant, si le « r » est trop grand, les petits détails seront perdus. Il y a des travaux en cours pour pouvoir faire le calcul automatique de « r ».

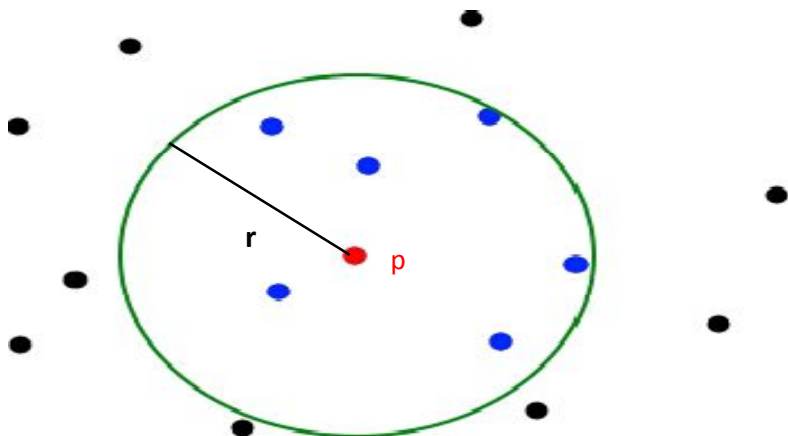


Figure II.1 : Voisinage d'un point P

2.2. Estimation de la normale dans un nuage de point :

La normale est une propriété importante d'une surface géométrique. Elle est largement utilisée dans de nombreux domaines tels que les applications d'infographie et d'autres effets visuels, même il existe de nombreuses stratégies basées sur la normale pour décrire la surface sous-jacent d'un nuage de point. L'orientation dominante en un point 3D est celle de la normale de la surface locale définie par le voisinage du point.

En géométrie, pour un point p_i situé sur une surface donnée, la normale n_i est le vecteur perpendiculaire à la surface de ce point. Un grand nombre d'algorithmes de nuage de point

Chapitre II Les descripteurs 3D

nécessite d'associer à tout point du nuage un vecteur unitaire supposé la normale à la surface échantillonnée.

Il y a plusieurs méthodes pour l'estimation de la normale mais dans notre travail la solution est réduite à une analyse des valeurs propres et vecteurs propres de la matrice variance-covariance, créé à partir des coordonnées des voisins les plus proches du point d'intégration. Plus précisément pour chaque point p_i , nous assemblons la matrice de covariance C comme suit :

$$C = \frac{1}{k} \sum_{i=1}^k (p_i - \bar{p}) * (p_i - \bar{p})^t \quad (\text{II.1})$$

Où k est le nombre de points considéré dans le quartier p_i et \bar{p} est la moyenne des voisins les plus proches.

La normale est le vecteur propre qui correspond à la plus petite valeur propre de la matrice de covariance [15].

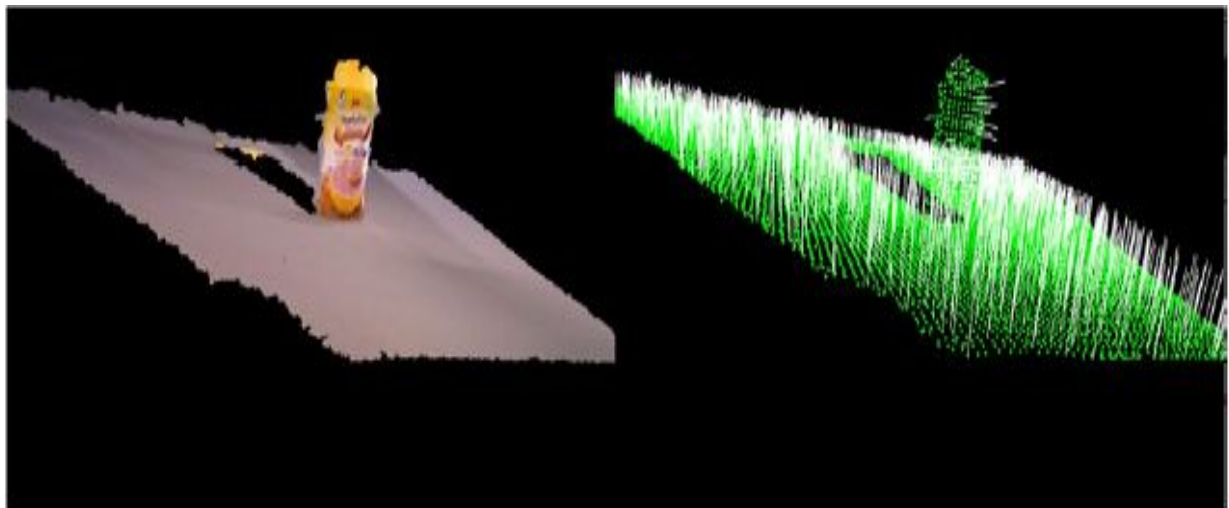


Figure II.2: nuage de point (gauche), estimation de la normale (droite).

3. Les descripteurs 3D :

3.1. Définition d'un descripteur 3D :

Un descripteur 3D est un outil essentiel dans la vision par ordinateur 3D pour comparer des surfaces et des ressemblances détectrices entre les points. Il est formé par un ensemble de

Chapitre II Les descripteurs 3D

valeurs caractéristiques, contenant des informations sur l'objet qu'il décrit, sous forme de vecteur ou matrice. Ces valeurs de descripteurs sont calculées à partir des coordonnées (x, y, z) de point p et celles de ses voisins. Un descripteur peut donc être représenté sous forme de vecteur par :

$$F = [f_1 f_2 f_3 \dots \dots f_d] \quad (\text{II.2})$$

Où d est la dimensionnalité du descripteur (nombre de dimension).

Idéalement, un descripteur possède les caractéristiques suivantes :

i) robuste, ii) stable, iii) unique et non ambigu, iv) invariant aux rotations aux translations et à la densité de point, v) discriminant.

Explication :

- i) La robustesse du descripteur signifie que ses valeurs ne changent pas ou très peu en présence de bruit de mesure.
- ii) La stabilité s'applique à l'extraction des points d'intérêt uniquement. elle signifie que la localisation des points sélectionnés ne varie pas ou très peu en présence de bruit.
- iii) L'unicité et le non ambiguïté sont importants afin que le calcul d'un descripteur, dans les mêmes conditions, mène toujours au même résultat.
- iv) L'invariance aux rotations et aux translations implique que le descripteur d'un point est le même peu importe le point de vue. Il s'agit d'une caractéristique primordiale pour qu'une même région soit reconnue entre des nuages de points acquis à des endroits différents.
- v) Le pouvoir discriminant est ce qui rend possible la distinction des descripteurs de points situés sur des surfaces. Par exemples, les valeurs calculées pour un point situé sur une surface plane ne devraient pas être les mêmes que celles obtenues pour un point situé sur un coin du mur. Cette caractéristique témoigne donc de la capacité à décrire avec précision la géométrie locale autour d'un point [14].

De nombreux descripteurs 3D ont été mis en œuvre dans PCL. Certains sont des extensions de descripteurs 2D, tandis que d'autres ont été spécifiquement développés pour caractériser des nuages de points 3D. dans notre projet nous allons nous intéresser à ces derniers.

Chapitre II Les descripteurs 3D

Les descripteurs peuvent être regroupés dans deux grandes catégories : les descripteurs locaux et les descripteurs globaux.

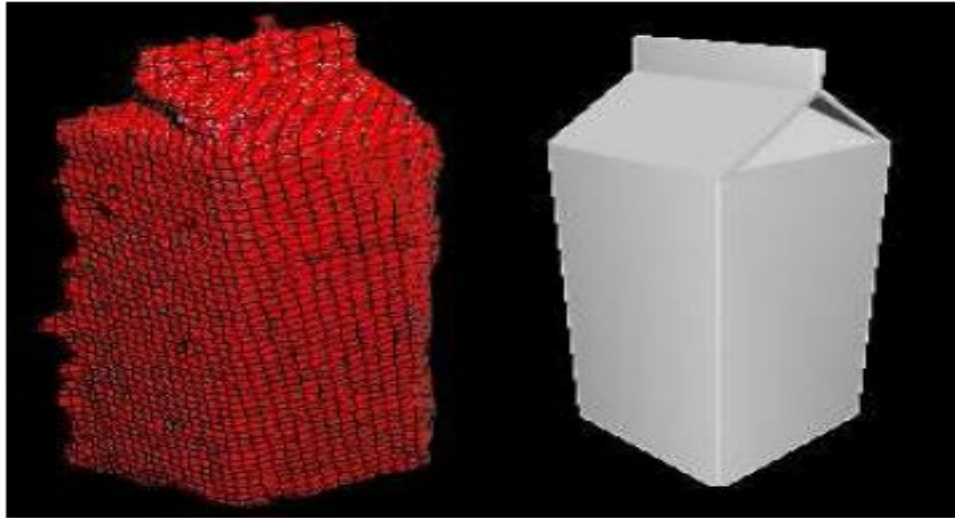


Figure II.3 : un modèle de description locale (gauche) et un modèle de description global (droite).

3.2. Descripteurs locaux :

3.2.1. Définition :

Un descripteur local est un descripteur calculé pour des points individuels que nous donnerons comme entrée. Il n'a aucune notion de ce qu'est un objet. Il décrit simplement la façon dont la géométrie locale est autour de ce point.

Les descripteurs locaux disponibles dans PCL sont :

- PFH (Point Feature Histograms)
- FPFH (Fast Point Feature Histograms)
- RSD (Radius-based Descriptor Surface)
- 3DS (Shape Context)
- USC (Unique Shape Context)
- SHOT (Signature Of Histograms Of orientations)

Chapitre II Les descripteurs 3D

-Spin Image

-RIFT (Rotation Invariant Feature Transform)

-NARF (Normal Aligned Radial Feature)

-RoPS(Rotational Projection Statistics)

Nous allons voir deux descripteurs les plus utilisés à savoir le PFH et le FPFH :

3.2.2. Point Feature Histograms (PFH):

Le point feature histogramme(PFH) est un descripteur qui donne un grand nombre d'informations sur la géométrie locale autour d'un point p . Son calcul est basé sur la combinaison de certaines relations géométriques entre les voisins les plus proches du point p [16].

Le PFH est discriminant, robuste au bruit, invariant aux rotations, aux translations et à la densité des points. Il encode les variations angulaires des normales entre les points du voisinage du point $\ll p \gg$ et les accumule dans un histogramme.

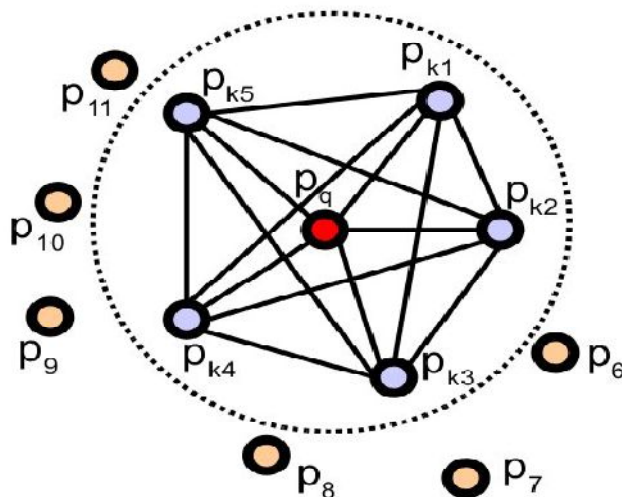


Figure II.4 : Région d'influence pour le calcul du PFH(le point rouge et ses k-voisins bleu sont entièrement interconnectés et délimité par une sphère de rayon r).

Le cercle pointillé représente la sphère délimitant le voisinage de point p . Les connexions entre les points marquent les points contribuant au calcul de descripteur.

Chapitre II Les descripteurs 3D

Le calcul de la PFH est comme suit :

- I. Pour chaque point p nous sélectionnons son voisinage qui est délimité par une sphère de rayon r .
- II. Pour chaque paire de points (p_i, p_j) ($i \neq j$) et leurs normales associées (n_i et n_j) dans le voisinage de p :
 - i. Nous choisissons Choisi p_i tel que $(\mathcal{A}_i, S) \leq (n_j, S)$, ou S est la droite reliant p_i et p_j (il est impératif que S aille dans une seule direction de p_i à p_j).
 - ii. Nous définissons un repère (U, V, W) ou U, V, W représentent un système de coordonnées de Darboux centré sur le point p_i en utilisant la procédure suivante :

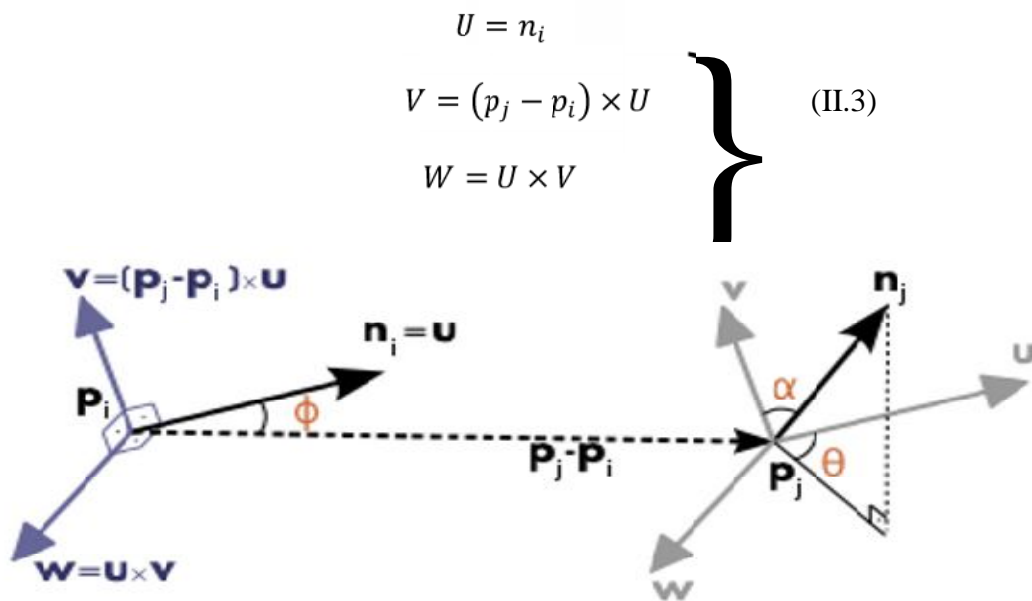


Figure II.5 : Illustration des variations angulaires calculées entre les normales de deux points lors du calcul d'un PFH.

- III. Nous utilisons l'équation (II.3) pour calculer les variations angulaires des normales [17] comme suit :

$$\alpha = V \cdot n_i$$

$$\varphi = (u \cdot (p_j - p_i)) / d \quad (II.4)$$

$$\theta = \arctan (W \cdot n_j, U \cdot n_j)$$

Où d est la distance entre les point p_i et p_j $d = \|p_i - p_j\|$ (II.5)

La distance d change avec le point de vue.

Chapitre II Les descripteurs 3D

Les angles α , φ , θ et la distance d sont calculés pour toutes les paires dans le k -voisinage du point p .

IV. Construire un histogramme à « b^d » nombre de bacs, où b est le nombre de bacs (c'est-à-dire les intervalles de subdivision dans plage de valeur d'une entité) et d le nombre de fonctionssélectionnées (comme Aldoma et al. proposé dans [18] : $b^d = 5^3 = 125$ cases), donc la dimensionnalité vaut 125.

Remarque :

Le PFH donne des résultats précis, mais il est trop coûteux en calcul pour être effectué en temps réel. Pour cela un descripteur dérivé plus rapide a été créé, il est nommé FPFH.

3.2.3. Fast Point Feature Histograms (FPFH) :

FPFH est une extension de PFH qui permet de donner des calculs moins élevés et conduit à un algorithme plus efficace. En plus il réduit le temps de calcul et conserve la majeure partie de la puissance descriptive de PFH. Il permet aussi de réduire la complexité computationnelle de l'algorithme de PFH de $O(n * k^2)$ à $O(n * k)$, où n le nombre de point dans le nuage et k le nombre de voisin de chacun des points [29]. En effet le FPFH est censé être utilisable dans des applications en temps réel. Il est conçu pour être robuste aux bruits de surface, et aux informations manquantes pour fonctionner de manière fiable.

Les principales différences entre les formulations PFH et FPFH sont résumées ci-dessous :

- Le choix des paires de points utilisées pour le calcul des descripteurs.
- Le FPFH ne fait pas interconnecter tous les voisins de point p .
- Le PFH modélise une surface précisément déterminée autour de point p délimité par le rayon r , tandis que le FPFH comprend des paires de points supplémentaires en dehors du rayon r .
- Le FPFH combine les valeurs de SPFH.
- La dimension de FPFH est plus petite que celle de PFH

Chapitre II Les descripteurs 3D

Le calcul de FPFH s'effectue sur l'ensemble d'un nuage de points comme suit :

- Calculer, pour chaque paire de point $\{p, p_k\}$ (point p et ses voisins p_k), les variations angulaires des normales pour établir des relations entre paires de points selon l'équation (II.4), en ajoutant un poids spécifique en fonction de la distance entre le point et chaque voisin. l'histogramme de point p obtenu est un résultat intermédiaire appelé un Simplified Point Feature Histogram (SPFH).

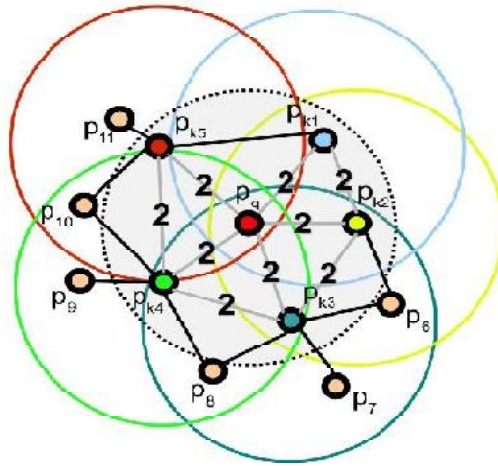
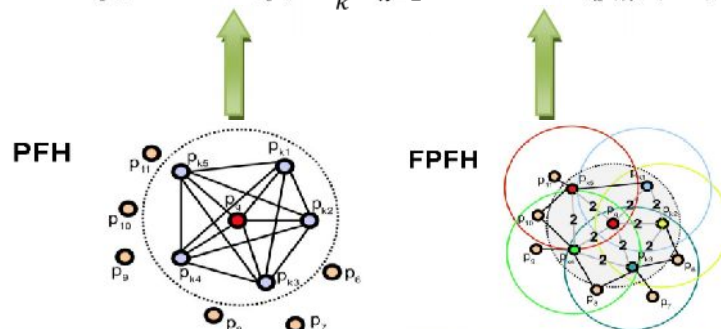


Figure II.6 : région d'influence pour le calcul de FPFH

La figure II.6 illustre la région d'influence pour le calcul du FPFH d'un point p . Le cercle grisé représente la sphère délimitant le voisinage de p . Les autres cercles de couleur représentent celles des voisins p_k [14].

- Dans la deuxième étape, pour chaque point p , nous recalculons les valeurs SPFH des voisins p_k , nous utilisons ces derniers pour calculer un histogramme final pour tout point p en additionnant à son SPFH celui de chacun de ses voisins p_k . Les SPFH des voisins sont pondérés en fonction de leur distance par rapport à p :

$$FPFH(p) = SPFH(p) + \frac{1}{k} \sum_{k=1}^k w_k \times SPFH(p_k) \quad (II.6)$$



Chapitre II Les descripteurs 3D

Où le poids w_k représente la distance entre le point et un point voisin p_k , $w_k = 1 / \|p - p_k\|$, et k le nombre de voisin de p .

Créant ainsi le FPFH, certaines des paires de valeurs seront comptées 2 fois.

La dimension de l'histogramme est plus petite que celle de PFH. Elle est définie par la relation $b \times 3$, où $b = 11$, comme l'on propose les auteurs dans [17], et donc la dimensionnalité est de $11 \times 3 = 33$.

3.3. Descripteurs globaux :

3.3.1. Définition :

Un descripteur global résume le contenu de l'image ou de la scène en un seul vecteur. Il a l'avantage de regrouper une grande quantité d'information en seulement une petite quantité de données. Il n'est pas calculé pour des points individuels, mais pour tout un ensemble qui représente un objet.

Ces type de descripteurs s'est révèlés être largement utilisés pour différentes tâches, en raison de son faible coût de calcul.

Les descripteurs globaux disponibles en PCL sont :

- VFH (viewpoint feature histograms)
- CVFH (clustered viewpoint feature histograms)
- ESF (Ensemble of shape functions)
- GFPH (global fast point feature histograms)
- GRSD (Global Radius-based Surface Descriptor)

Nous allons voir ici deux des descripteurs les plus utilisés, à savoir VFH et le CVFH :

3.3.2. ViewpointFeatureHistograms(VFH) :

L'histogramme de point de vue est un descripteur pour les nuages de points 3D qui encode la géométrie et ajoute des informations sur le point de vue variant. Il a été créé à base

Chapitre II Les descripteurs 3D

de FPFH qui est invariant à la pose de l'objet. En outre, Le VFH contient deux composantes : une composante représente le point de vue et une composante FPFH étendu [19].

Pour calculer la première composante, il faut d'abord trouver le barycentre de l'objet, qui est le point qui résulte d'une moyenne des coordonnées X, Y, Z de tous les points. Ensuite, le vecteur entre le point de vue (la position du capteur) et ce barycentre est calculé et normalisé. Enfin, pour tous les points de la grappe, l'angle entre ce vecteur et leur normale est calculé, et le résultat est mis en cellule dans un histogramme. Le vecteur est traduit à chaque point lors du calcul de l'angle, car il rend l'échelle de descripteur invariant.

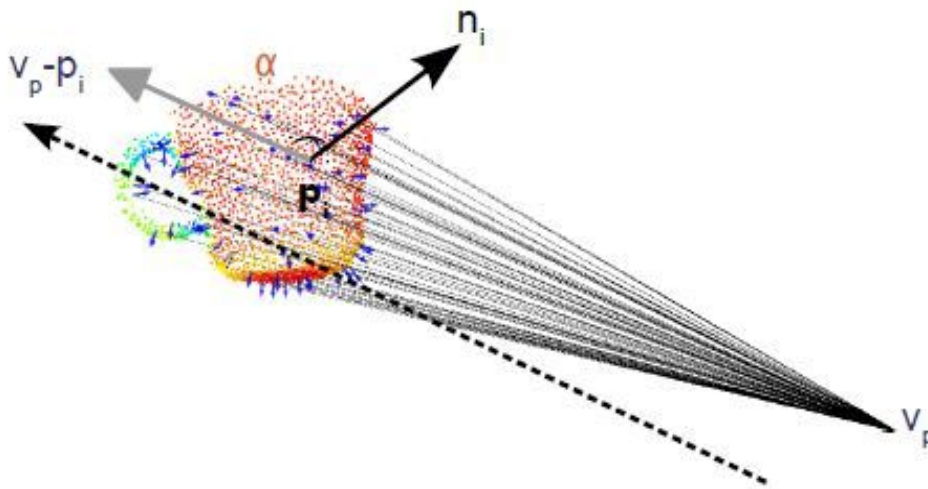


Figure II.7: barycentre et vecteur normalisé.

La deuxième composante est calculée comme le FPFH avec quelques différences : il est calculé uniquement pour le barycentre, en utilisant le vecteur de direction de point de vue calculé comme normale et le reste des points de la grappe en tant que voisins.

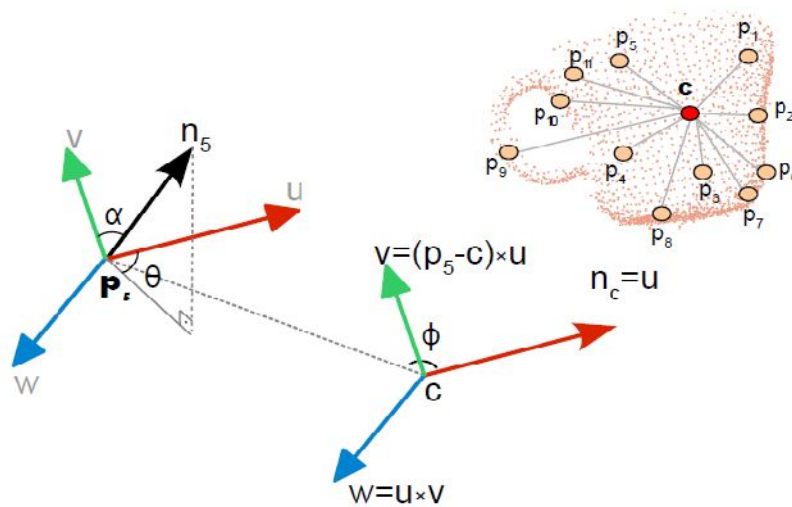


Figure II.8 : calcul des composantes angulaires pour VFH

Les 4 histogrammes résultant (un pour le composant de point de vue, trois pour le composant FPFH étendu) sont concaténés pour construire le descripteur final VFH. Par défaut, les bacs sont normalisés en utilisant le nombre total de points dans le cluster. Cela rend le descripteur VFH invariant à l'échelle [20].

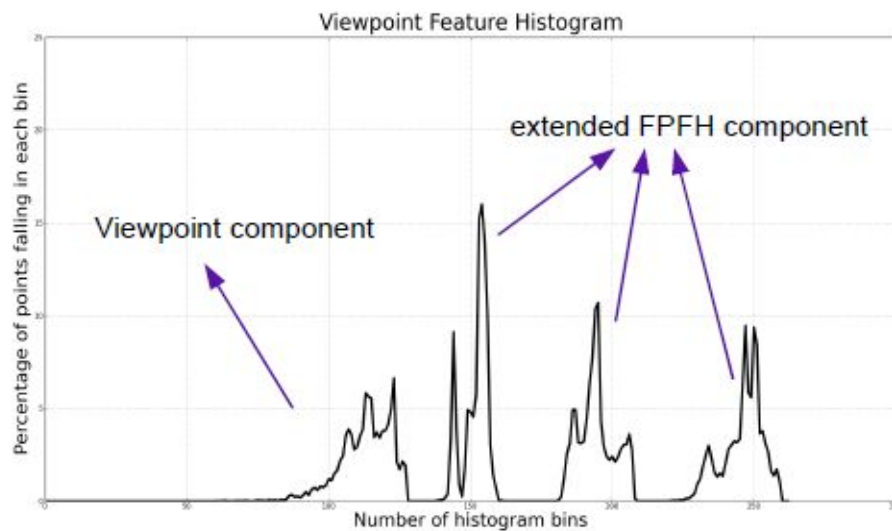


Figure II.9: histogramme du VFH

La dimensionnalité du descripteur VFH est de $4 \times 45 + 128 = 308$, la dimensionnalité est divisée en deux parties (deux bacs de taille différente) car ce descripteur a deux composants. La taille b du bac est de 45 pour la première et 128 pour la seconde.

La valeur 4 représente le nombre de composants du cadre de Darboux et la complémentaire information sur la distance [21].

3.3.3. Clustered Viewpoint Feature Histograms (CVFH):

CVFH est une extension du descripteur VFH. Ce dernier, il n'est pas robuste à l'occlusion et aux erreurs de mesure, ce qui fait qu'il y'a des pertes de points dans le cluster de l'objet. Cela induit des erreurs dans le calcul du barycentre, et qui a pour effet de modifier le descripteur final. A cause de cela, le CVFH a été introduit.

L'idée est très simple : au lieu de calculer un seul histogramme VFH pour l'ensemble du cluster, l'objet est d'abord divisé en régions lisses et stables en utilisant la segmentation « region-Growing », qui impose plusieurs contraintes dans les distances et les différences de normale des points appartenant à chaque région, Ensuite, un VFH est calculé pour chaque région. Grace à cela, un objet peut être trouvé dans une scène, dès l'or qu'au moins l'une de ses région est entièrement visible [20].

Le CVFH est obtenue en utilisant les étapes suivantes :

- Déterminer l'ensemble S des régions lisses et stables.
- Pour chaque $s_i \in S$, trouver le centroïde (c) et sa normale (n_c).
- Construire un cadre de référence local (u_i, v_i, w_i) comme dans le VFH.
- Trouver les histogrammes des angles ($\alpha, \phi, \theta, \beta$) comme dans le VFH (les 3 premiers codées comme 45-bin Histogrammes et β coder comme un histogramme 128-bin histogramme).
- Trouver le composant de distribution Shape (SDC) en tant que

$$SDC = \frac{(c-p_i)^2}{\max\{(c-p_i)^2\}}, i = 1, \dots, |S| \text{ (II.7)}$$

Le Shape distribution component (SDC) est également calculé et inclus, il code des informations des points au tour du barycentre de la région, et mesure les distances. La SDC permet de différencier des objets présentant des caractéristiques similaires (taille et distribution de normale).

Le CVFH est donnée par des histogrammes concaténés ($\alpha, \phi, \theta, SDC, \beta$) qui ont une dimensionnalité de 308 [21].

Les auteurs ont proposé de rejeter l'étape histogramme de normalisation qui est effectué dans le VFH. Cela a pour effet de rendre le descripteur dépendant de l'échelle, donc

Chapitre II Les descripteurs 3D

un objet d'une certaine taille ne serait pas correspondre à une copie plus ou moins important de lui-même. Il rend également CVFH plus robuste à l'occlusion.

Le CVFH est invariant à l'angle de roulis de la caméra, comme la plupart des descripteurs globaux. Il en est ainsi parce que les rotations autour de cet axe de la caméra ne changent pas la géométrie observable à partir de lequel les descripteurs sont calculés.

Ce tableau résume quelques points importants sur les descripteurs présents dans ce chapitre :

Descripteur	n point	Point size	Normale
PFH	N	125	Y
FPFH	N	33	Y
VFH	1	308	Y
CVFH	m n	308	Y

Tableau I.1 : Résumé de quelque point important des descripteurs présenté dans ce chapitre.

-La première colonne contient les différents descripteurs présents dans ce chapitre.

-La deuxième colonne contient le nombre de points générés par chaque descripteur, donnée un nuage de point d'entrée avec n points.

-La troisième colonne montre la dimensionnalité de chacun des points générés.

-La dernière colonne indique si l'algorithme requiert le calcul des normales de surface à chaque point.

Pour une description plus détaillée des différents descripteurs de nuage de points 3D, nous invitons le lecteur intéressé à consulter [17, 20, 21].

4. Conclusion :

Dans ce chapitre nous avons vu comment les travaux de recherche ont développé le domaine de reconnaissance d'objets 3D, par l'utilisation des descripteurs. Ces derniers donnent la possibilité de décrire la géométrie autour d'un point et donnent un ensemble de valeurs caractéristiques contenant des informations sur l'objet qu'il décrit. Chaque descripteur a son propre algorithme de calcul, avec ses propres paramètres. Par conséquent, cela donne de différentes valeurs et de différentes caractéristiques. Idéalement, un descripteur possède les caractéristiques suivantes : robuste, stable, unique et non ambigu, invariant aux rotations aux translations et à la densité du point, discriminant. Il y a aussi d'autres caractéristiques intéressantes d'un descripteur : sa rapidité de calcul et sa faible dimensionnalité.

Il existe plusieurs descripteurs 3D mis en œuvre dans PCL. Ils peuvent être classés en deux catégories principales : globales et locales. Parmi les descripteurs locaux, le descripteur PFH qui est très discriminant, robuste au bruit et invariant aux rotations et aux translations et donne des résultats très précis, mais il est trop coûteux en calcul pour être effectué en temps réel. C'est d'ailleurs pour cela que le descripteur FPFH a été créé : pour être discriminant aussi et plus rapide que le PFH. Pour les descripteurs globaux, il existe le descripteur VFH qui est variant à la pose de l'objet contrairement au FPFH, mais il n'est pas robuste à l'occlusion et aux erreurs de mesure, ce qui a conduit au développement de CVFH. Ce dernier il est robuste à l'occlusion et aux erreurs de mesure. En plus, il est variant par rapport aux changements d'échelle.

Nous ne pouvons pas désigner la meilleure catégorie de descripteurs, car les descripteurs calculés pour l'objet entier « descripteurs globaux » permettent de réduire le temps de calcul pour l'utiliser dans des applications en temps réel mais ne sont pas efficaces pour certaines applications. A l'inverse, les descripteurs calculés pour des points individuels « descripteurs locaux » sont efficaces, mais coûteux et cela limite leur application pour le traitement en temps-réel. Par exemple, la reconnaissance d'objet avec un robot mobile. Donc le choix du descripteur, qu'il soit local ou global, dépend de l'application visée.

Chapitre III :

La classification automatique d'objets 3D

1. Introduction :

« Le seul moyen de faire une méthode instructive et naturelle, est de mettre ensemble les choses qui se ressemblent et de séparer celles qui diffèrent les unes des autres. »M. Georges Buffon, Histoire naturelle, 1749.

Depuis l'aube des temps, l'homme pratique la classification dans sa vie quotidienne, quand il essaie de répondre aux problèmes et questions sur la catégorie des objets, c'est-à-dire d'affectation d'objets à leur classe (en observant leurs formats, couleurs, tailles...etc.). Il est clair que le processus générale de la classification dans le domaine informatique essaie de l'appliquer sur des données numériques (points, images, sons,...etc.), cela n'échappe pas à la règle imposée par ce célèbre naturaliste et écrivain Georges Buffon, et que le travail général des méthodes de classification, depuis 1749, consiste à imiter et automatiser ce principe en utilisant et inventant des moyens adéquats (matériaux, calculateurs, et des théories classificatoires...etc.)

Dans le cadre des problèmes de classification, nous disposons d'un ensemble de données qui reprend une collection d'objet non étiqueté. Les classes sont encore inexistantes. L'objectif alors est d'obtenir des classes d'objets homogènes en favorisant l'hétérogénéité entre ses différentes classes [22].

Les méthodes de classification sont divisées en deux groupes :La classification non supervisée, et la classification supervisée. De même certains problèmes de classification nécessitent de combiner les deux types d'apprentissage (supervisée et non supervisée), cette méthode s'appelle la méthode semi supervisée.

Dans ce chapitre, nous allons donc, en premier lieu, définir la classification et voir les critères pour une bonne classification. Ensuite nous allons présenter les méthodes de classifications et détailler une approche de la méthode de classification non supervisée "Algorithme des k-means", qui est l'algorithme que nous allons utiliser dans notre application. Puis, en deuxième lieu, nous allons présenter les différentes étapes de notre application telle que l'acquisition, le traitement, calcul de VFH...

2. Définition :

La classification est une discipline reliée de près ou de loin à plusieurs domaines. Elle est connue aussi sous des noms variés (classification, clustering, segmentation) selon les objets qu'elle traite et les objectifs qu'elle vise à atteindre.

Pour attribuer une définition au terme « classification », il faudrait d'abord définir ses racines, qui viennent du verbe « classer » qui désigne plus une action qu'un domaine, ou plutôt une série de méthodes qu'une théorie unifiée.

En mathématique, nous appelons la classification, la catégorisation algorithmique d'objets. Elle consiste à attribuer une classe ou catégorie à chaque objet à classer, en se basant sur des données statiques. Elle fait couramment appel aux méthodes d'apprentissage. Elle est largement utilisée en reconnaissance des formes.

3. Critères pour une bonne classification :

Une bonne classification devrait satisfaire ces différents critères :

▪ Validité :

Les objets qui appartiennent à la même classe doivent être semblables, et les objets de différentes classes doivent être différents.

▪ Interopérabilité :

Les classes doivent avoir une interprétation substantive c'est-à-dire qu'il est possible de donner des noms aux classes.

▪ Stabilité :

Les classes doivent être stables, cela veut dire que les petites modifications dans les données et dans les méthodes ne doivent pas changer les résultats.

4. Les étapes de classification :

1. Choix des données.
2. Calcul des similarités entre les n individus à partir des données initiales.
3. Choix d'un algorithme de classification et exécution.
4. L'interprétation des résultats.

5. Méthodes de classifications :

Les méthodes de classifications peuvent être séparées en deux grandes catégories : les méthodes de classifications supervisées et les méthodes de classification non supervisée.

5.1. Classification supervisée :

C'est l'ensemble des techniques qui visent à deviner l'appartenance d'un objet à une classe en s'aidant des valeurs qu'il prend. Elle construit un modèle représentatif d'un certain nombre de données organisées en classes que nous appellerons généralement le corps-d'apprentissage, puis d'utiliser ce modèle afin de classer de nouvelles données, c'est à dire prédire leur classe au vu de leur caractéristiques (appelées paramètres).

La plupart des algorithmes d'apprentissage supervisés tentent de trouver un modèle (une fonction mathématique) qui explique le lien entre des données d'entrée et les classes de sortie. Ces jeux d'exemples sont donc utilisés par des algorithmes.

Il existe de nombreuses méthodes d'apprentissage supervisé [23], parmi elles la méthode des K plus proches voisins et la méthode des Réseau de neurones.

5.2. Classification non supervisée :

L'objectif de ces méthodes est de grouper les individus en un nombre restreint de classes homogènes sans connaissances à priori.

L'apprentissage non supervisé consiste à inférer des connaissances sur des classes sur la seule base des échantillons d'apprentissage, et sans savoir a priori à quelle classe ils appartiennent [23].

5.2.1. Principe de Classification non supervisé :

Contrairement aux méthodes de classification supervisées, nous ne possédons pas des connaissances à priori sur les classes prédéfinies des éléments. Donc la division des objets dans les différents groupes (clusters) se fait en se basant sur le calcul de similarité entre les éléments.

Alors que l'objectif des méthodes du Clustering est de grouper des éléments proches dans un même groupe, de manière à ce que deux données d'un même groupe soient le plus similaires possible et que deux éléments de deux groupes différents soient le plus dissemblables possible [24].

Mathématiquement, nous avons un ensemble X de N données décrites chacune par leur P attribut. Donc le clustering consiste à créer une partition ou une décomposition de cet ensemble en sous parties (clusters) telle que :

- les données appartenant au même groupe se ressemblent.
- les données appartenant à deux groupes différents soient peu ressemblantes.

Remarque :

Un cluster est une collection d'objets qui sont « similaires » entre eux et qui sont « dissemblables » par rapport aux objets appartenant à d'autres groupes.

5.2.2. Méthodes de Classification non supervisée :

Nous distinguons deux types de Clustering : les approches de classification hiérarchiques et les approches de classification non hiérarchiques.

a. Méthodes hiérarchiques :

La classification hiérarchique consiste à effectuer une suite de regroupements en classes, de moins en moins fines, en agrégeant à chaque étape les objets ou les groupes d'objets les plus proches. Elle fournit ainsi un ensemble de partitions de l'ensemble d'objets [25]. Cette approche utilise la notion de distance, qui permet de refléter l'homogénéité des classes. Ainsi, nous considérons qu'un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres.

b. Méthodes non hiérarchiques :

Ce sont des méthodes qui produisent directement une partition en un nombre fixé de classe [26].

Regrouper n individus en k classes de telle sorte que les individus d'une même classe soient le plus semblables possible et que les classes soient bien séparées.

Parmi ces méthodes, nous retrouvons :

i. Méthode des k-means :

L'algorithme k-means mis au point par McQueen, en 1967, est un des plus simples algorithmes d'apprentissage non supervisé, appelé algorithme des centres mobiles [27], le attribue chaque objet à un cluster dont le centre est le plus proche. Le centre est la moyenne de tous les points dans le cluster. Ses coordonnées sont la moyenne arithmétique pour chaque dimension séparément de tous les points dans le cluster, c'est-à-dire chaque cluster est représenté par son centre de gravité.

➤ **Avantage des K-means :**

-L'avantage de ces algorithmes est avant tout leur grande simplicité.

-Applicable à des données de grandes tailles.

➤ **Inconvénients des K-means :**

-Le nombre de classes doit être fixé au départ.

-Ne détecte pas les données bruitées.

-Le résultat dépend du tirage initial des centres des classes.

❖ **Algorithme des k-means :****Entrée**

Nombre de classe souhaité, notée par n

Sortie

Une partition de n classe $\{C_1, C_2, \dots, C_n\}$

Début

Initialisation aléatoire des centres C_n ;

Initialisation $Som(n) = 0$

Répéter

Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche :

$e_{ij} \in C_n$ Si la distance $min = d(e_{ij}, C_n)$

$$sum(n) = som(n) + e_{ij}$$

$k_n ++$

Représentation : calcul les nouveaux centres :

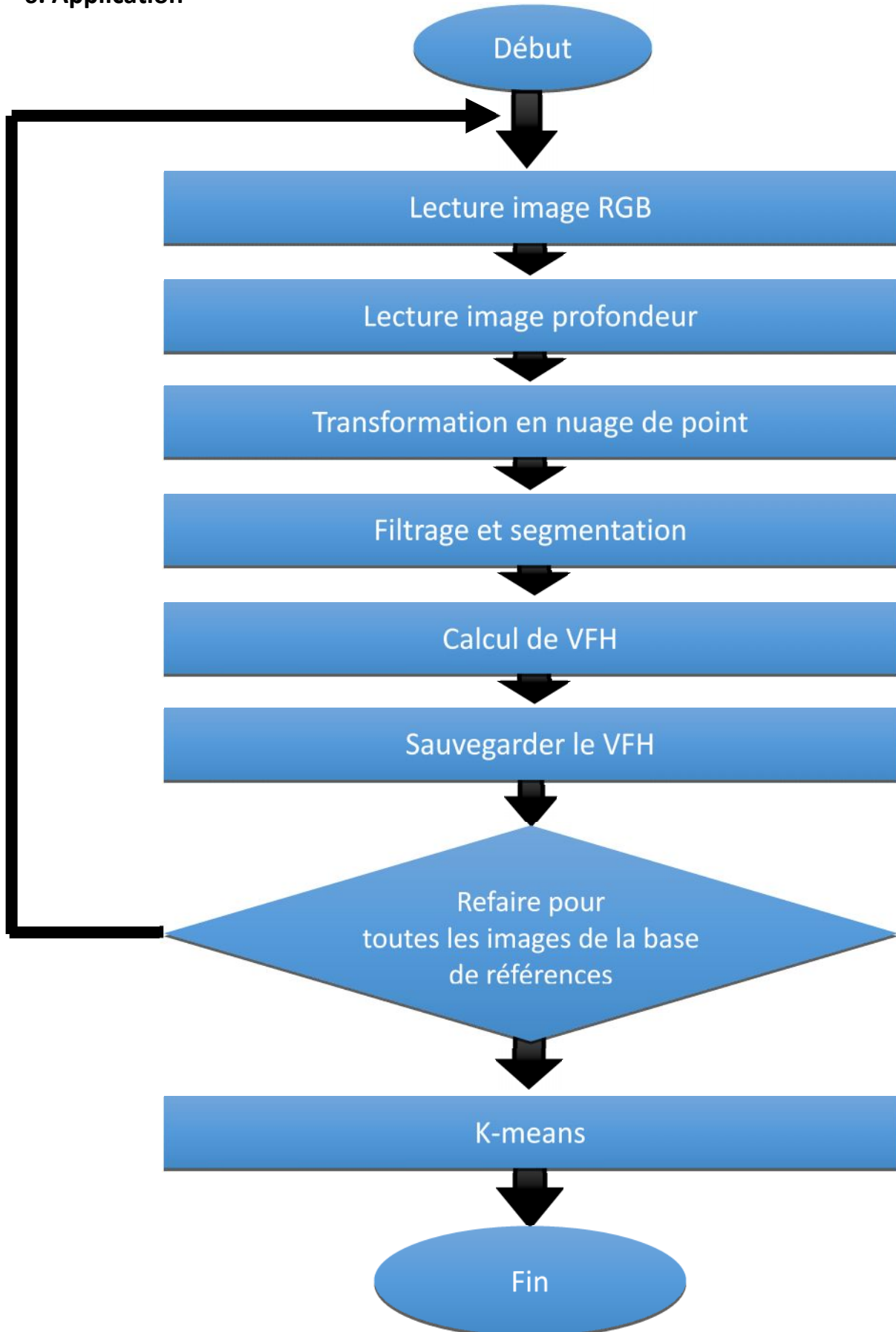
$$C_n = \frac{Som(n)}{K_n}$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

$$C_{n+1} - C_n = 0$$

Fin

6. Application



6.1. Extraction des paramètres 3D :

6.1.1. Acquisition :

Dans ce travail nous allons utiliser le système d'acquisition Kinect de Microsoft, qui va nous permettre de récupérer des images de profondeur et de couleur (RGB-D), comme illustre la figure III.1.

- **Transformation en nuage de point :**

Avec les informations acquises par la Kinect, Nous pouvons créer pour chaque vue le nuage de point correspondant, comme nous l'avons cité dans le chapitre I.

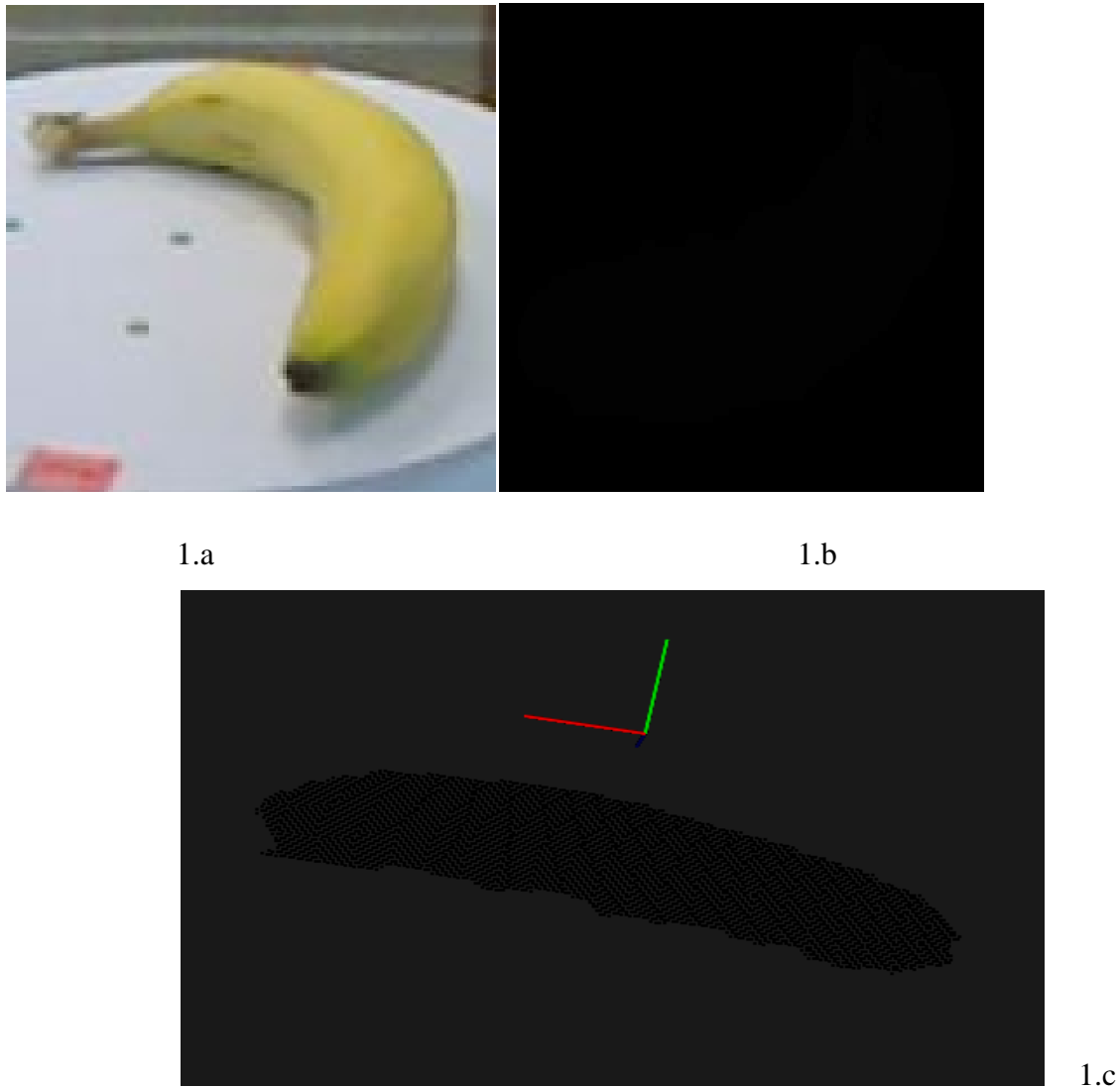


Figure III.1 : 1.a image couleur, 1.b : image de profondeur, 1.c : nuage de point.

6.1.2 Traitement :

Cette étape a pour but de préparer, l'image (éliminer le bruit, séparer l'objet du fond) pour l'utiliser dans la prochaine étape. Elle est divisée en deux parties :le filtrage et la segmentation.

- **Filtrage :**

Un nuage de point brut, tel qu'il a été renvoyé par le capteur, est généralement un nuage bruité. Causé par la faible résolution de la caméra, mauvais éclairage..., le bruit conduit à des mesures incorrectes, ce qui fait que le filtrage est une étape importante dans la chaîne de reconnaissance de forme.

Pour éliminer le bruit deux filtrages sont possible, le premier, en utilisant des filtres matériels (placer directement sur le capteur), le deuxième, en utilisant des filtres logiciels (programme informatique). Le premier est plus bénéfique en matière de gain de temps.

- **Segmentation :**

Fondamentalement, la segmentation est un processus qui consiste à découper une image en régions connexes présentant une homogénéité selon un certain critère (couleur, texture, niveau de gris, indice,..) [28], ou plus simplement, c'est la méthode qui permet de récupérer l'objet qui présente un intérêt pour nous à partir d'une scène encombrée, autrement dit, séparer l'objet du fond.

**a****b**

Figure III.2 : Exemple d'image segmentées : (a) pomme, (b) banane

La classe bol contient 785 objets

La classe boite de céréale contient 592 objets

La classe tasse à café contient 982 objets

La classe assiette contient 887 objets

6.2.2. Matériel et logiciel utilisé :

- **Matériel :**

Pc :	DELL
Processeur :	Intel Core i5 / 1.80Ghz
RAM:	4Go
- **Logiciel :**
 - Système d'exploitation UBUNTO 14.04 (de Linux).
 - Langage de programmation C++.
 - Système d'exploitation Windows 64 bits.
 - Langage de programmation Matlab 2016.
 - La bibliothèque : Point Cloud Library

6.2.3. Résultats

La figure III.4 représente la projection des six classes :

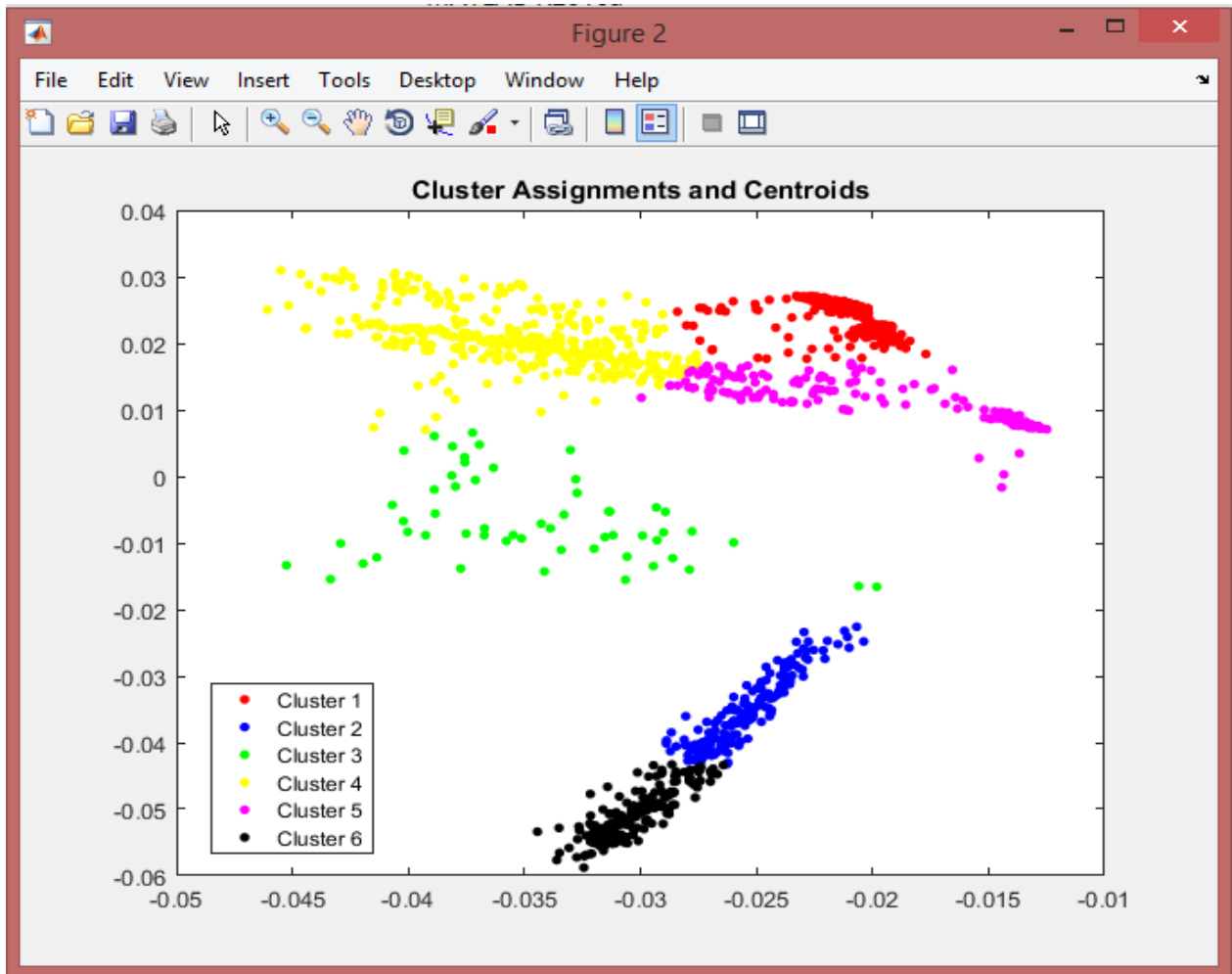


Figure III.4 : la projection des six classes.

Après étiquetage nous avons obtenu la matrice de confusion représentée par la table III.1 :

	Pomme	Banane	Bol	Boite de céréale	Tasse à café	Assiette
Pomme	503	11	60	0	26	25
banane	4	487	5	40	22	11
Bol	10	6	615	3	52	99
Boite de céréale	23	9	47	434	39	40
Tasse à café	98	29	60	83	628	84
Assiette	5	1	19	4	6	852

Tableau III.2 : Matrice de confusion

- **La précision :**

Pommes : 80 % de bonne reconnaissance

Bananes : 85 % de bonne reconnaissance

Bol : 78 % de bonne reconnaissance

Boite de céréale : 73 % de bonne reconnaissance

Tasse à café : 64 % de bonne reconnaissance

Assiette : 96 % de bonne reconnaissance

Total : 79% de bonne reconnaissance.

7. Conclusion :

Nous avons présenté dans ce chapitre des généralités sur les conceptions de la classification et un aperçu sur ces principales méthodes, parmi celles ci les méthodes non supervisées, qui permettent d'obtenir des informations sans aucune connaissance préalable, ce qui n'est pas le cas des méthodes supervisées.

Plusieurs méthodes sont proposées pour le problème de la classification non supervisée. Face à ces méthodes, les scientifiques sont souvent incapables de designer la meilleure d'entre elles, car chacune de ces méthodes possède ses points forts et ses points faibles. Les méthodes hiérarchiques sont utilisées en cas des données de petite taille, car la complexité est très élevée. Si des problèmes de temps d'exécution se posent, alors c'est les méthodes non hiérarchiques qui sont utilisées. Pour cela dans notre application nous nous sommes intéressés à une approche de cette dernière nommée la méthode des K-means .

Dans cette application, le test est effectué sur une base de données composé de 06 classes de 4 440 objets.

Les résultats obtenus sont globalement bons, puisque nous avons un taux de bonne reconnaissance global de 79%, la meilleure reconnaissance a été enregistrée sur les assiettes avec un taux de reconnaissance de 96 %, et une reconnaissance satisfaisante pour la pommes et les bananes, par contre la reconnaissance pour les bols, les boites de céréale et les tasses à café est un peut faible. Ces résultats nous permettent de comprendre que ces paramètres de descripteur VFH sont largement suffisants pour la reconnaissance de certains objets 3D.

Pour accélérer les résultats de cette application nous proposons d'utiliser une station de calcul plus puissante, et pour perfectionner notre travail nous proposons d'essayer d'autres descripteurs telle que le descripteur CVFH qui est robuste à l'occlusion, et de réduire la taille des descripteurs avec l'Analyse des Composantes Principale (l'ACP). Cette dernière permet, non seulement, un gain en temps de calcul et en espace mémoire, mais améliore la robustesse du descripteur au bruit.

Conclusion générale

Nous rappelons que l'objectif de ce projet de fin d'étude est le calcul des descripteurs 3D dans le but de la classification automatique d'objets 3D, qui est une phase de recherche qui fait appel à plusieurs domaines. Nous avons commencé par un bref état de l'art sur l'acquisition d'image 3D à savoir les méthodes actives et les méthodes passives, puis nous avons présenté le principe de fonctionnement et les différentes composantes de la caméra Kinectde Microsoft qui nous a permis d'avoir les données couleurs (RGB) et de profondeurs des images que nous avons utilisées dans notre application, ainsi que les différentes manières de représenter les images tridimensionnelles telle que la carte de profondeur et les nuages de points.

Ensuite, nous avons expliqué quelques notions de base indispensables sur les nuages de points telle que la notion de voisinage est le calcul de la normale. Nous avons aussi défini les descripteurs 3D, leurs types globaux ou locaux, leurs caractéristiques et le principe de fonctionnement de certains d'entre eux.

Finalement, nous avons présenté dans le troisième chapitre une généralité sur les conceptions de la classification et un aperçu sur ces principales méthodes, ainsi que la méthode des K-means que nous avons utilisée dans ce travail, suivie des différentes étapes de notre application telle que l'acquisition, le traitement, le calcul de descripteur VFH...etc.

Les tests de notre application sont effectués sur une base de données composée de 440 objets 3D, et les résultats ont donné un taux global de bonne reconnaissance de 79 % qui est satisfaisant.

Pour accélérer ce résultat, nous proposons d'utiliser une station de calcul plus puissante, et pour perfectionner notre travail, nous proposons d'utiliser d'autres descripteurs comme le descripteur CVFH, qui est robuste à l'occlusion et de réduire la taille des descripteurs avec l'Analyse des Composantes Principale (l'ACP). Cette dernière permet non seulement, un gain en temps de calcul et en espace mémoire, mais améliore la robustesse du descripteur au bruit.

Ce travail nous a donnée la chance d'utiliser un nouveau système d'exploitation nommé UBUNTO 14.04 (de linux) et le langage de programmation C++, que nous n'avons jamais utilisé durant notre parcours d'études. Il nous a également permis de découvrir le monde de la recherche, et d'enrichir nos connaissances dans le domaine de la vision par ordinateur 3D.

Références :

- [1] Thèse de Boubbaba ben AMOR,(contributions à la modélisation et à la reconnaissance Faciales 3D)page 9.
- [2] Abban Bajard . « Normalisation 3D de surface métalliques spéculaires par imagerie infrarouge » thèse doctorat Université de BOURGONE. Page 8 et 9.
- [3] Elodie Lucas « la stéréovision » article page 2
- [4] Thèse de Bastien Billiot (conception d'un dispositif d'acquisition d'images agronomique 3d en extérieur et développement des traitements associés pour la détection et la reconnaissance de plantes et de maladies).page 23
- [5] N. Payet and S.Todorovic «scene Shape from texture of objects». In IEEE Conf. on computer vision and pattern Recognition. 2011.pp.2017-2024
- [6] R.Zhang ,p.s. Tsai.J.E . Cryer.and M.shah, « Shape from Shading » . IEEE transactions on pattern Analysis and Machine Intellegence,Vol.21,no.8,pp.690-706.
- [7] Benjamin loriot, Automatisation de l'acquisition et des post-traitements en Numérisation 3D, Université de Bourgogne, thèse de doctorat soutenue mars 2009.
- [8] magister de Mansouri douel kafel « la comparaison des images 3D » Univercité d'oran 2011 .chapitre 1
- [9] Robinson,A.Abboul,L. and Rodrigues ,M. « Methods for indexing stripes in uncoded structured light scanning systéms». Journal of winter school of computer Graphics, 2004,12(1-3).
- [10] Bertrand pecuchet, « Kinect fir Xbox 360 » Université Lille 1, 2012/2013.
- [11] J.Han,L.Shao,D.xu,and j.Shotton. Enhanced computer vision with Microsoft kinect sensor : A review. IEEE trans, cybernetics, 43(5):1318-1334, 2013.
- [12] K. khoshelham and S.O. Elberink Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors, 12(2):1437-1454, 2012.

- [13] Sébastien Poulmane, Guillaume Dauster, Marc Amberg «Reconstruction de scene réaliste via kinect, amelioration du rendu des textures plaques dans une scène observée une kinect et reconstruire avec pcl » Université Lille 1. 2011/2012. Page 10 et 11.
- [14] Maxime Latulippe «calage robuste et accéléré de nuage de point en environnement naturel via l'apprentissage automatique ». Université LAVAL, Québec, Canada, 2013.
- [15] <http://pointclouds.org/documentation/Estimating> surface Normals in a Point Cloud.
- [16] R. Russu, N. Blodow, Z. Marton, and M. Beetz «Aligning point cloud views using persistent feature histograms» in International conference on Intelligent Robots and systems (IROS), Nice, France, September 26-2-2008.
- [17] R.B. Rusu, N. Blodou, and M. Beetz« fast point feature histograms (FPFH) for 3D registration» in proceedings of the 2009 IEEE. International conference on robotics and automation ser.ICRA09.
- [18] Aldoma A, Marton ZC, Tombari F, Wohlkinger W, Potthast C, Zeisl B, Rusu RB, Gedikli S, VinczeM (2012) Tutorial: Point cloud library: Three-dimensional object recognition and 6d of pose estimation. IEEE Robotics & Automation Magazine, 1070 (9932/12), pp 80-91.
- [19] R.B. Rusu, G. Bradski, R. Thibaut, and J.Hsu«Fast 3D Recognition and pose using the viewpoint Feature histograms» in proceedings of the 23 rd IEEE international conference on intelligent Robots and systems(Iros), Taipei, Taiwan, October 2010.
- [20] [//robotica.unileon.es/mediawiki/index.php/PCL/openNI_tutorial_4:3D_object_recognition](http://robotica.unileon.es/mediawiki/index.php/PCL/openNI_tutorial_4:3D_object_recognition) (descriptors).
- [21] Luis A. Alexandre"3D descripteur for object and category Recognition: a comparative Evaluation".
- [22] Mounzer BOUBOU : contribution aux méthodes de classification non supervisée via des approches prétopologiques et d'agrégations d'opinion , thèse de doctorat, université Claude Bernard- Lyon, 2007.

- [23] Belhabib Abdelkader, Lagha Omar : Développement d'une application à base de l'algorithme de classification K-means , Mémoire de fin d'étude pour l'obtention du diploma de licence en Informatique, Université Abou bakr belkaid, Tlemcen, 2012.
- [24] [Hartigans, 1975] : J. Hartigans. Clustering algorithms. John Wiley and Sons, Inc.,1975.
- [25] [Dempster and al, 1975]: A.P Dempster, N.M Laird, and D.B Rubin Maximum- vraisemblance from incomplete data via the algorithm , J.Royal Statist, society,B39, pp 1-38,1997.
- [26] [Bezdek, 1981] :J.C Bezdek (1981): pattern recognition with fuzzy objective function algorithms , plenum press, New York 1981.
- [27] [Guillaume & Hervé, 2002]: LECOINTRE Guillaume et le Guyader Hervé, Classification phylogénétique du vivant , 2e edition, 2002, Belin, Paris.
- [28] J.-P. Cocquerez et S. Philipp, « Analyse d'image : Filtrage et segmentation », Edition MASSON. Chapitre 4.

Annexe :

Point Cloud Library :

Historique de la bibliothèque Point Cloud Library :

PCL a été développés au cours de doctorat du Dr Radu Bogdon Rusu et Muenchen, à l'Université de Technicien, en Allemagne. Puis raffinés à Willow Garage à la fin de 2009. Le travail de Rusu était de créer une infrastructure commune pour la recherche et les applications de traitement 3D. A partir de mars 2010, L'idée d'avoir un ensemble de bibliothèques modulaires a commencé à prendre forme et certain nombre d'outils ont été écrits dans le cadre du système d'exploitation Robot(ROS) pour aider les robot à manipuler et à naviguer dans des environnements 3D complexes. Le travail a continué, mais ce n'est qu'en novembre 2010, lorsqu'une décision a été prise de commencer à créer un nouveau projet autonome appelé PCL qui profiterait à la plus grande communauté 3D dans son ensemble. À la fin de mars 2011, PCL a pris ses premières mesures et a lancé comme projet distinct avec son propre site web (www.pointclouds.org).

Présentation de la bibliothèque point Cloud Library :

Avec la démocratisation des scanners laser de nouvelle génération, de plus en plus de personnes de tout bord acquièrent des données 3D en temps réel et à moindre coût. Il est souhaitable que des outils de traitements et d'algorithmes 3D ainsi que des normes soient partagé par le plus grand nombre. C'est exactement le rôle que joue la bibliothèque Point Cloud Library (PCL). Elle vise à fournir une collection d'algorithmes et d'outils pour faciliter les recherches dans le domaine de traitement de nuage de point 3D.

Cette bibliothèque contient des algorithmes pour la reconstruction de surface, la segmentation, le filtrage, la classification, la reconnaissance des objets par rapport à leurs géométries,... etc.

PCL est un projet open-source sous licence BSD, il est gratuit pour la commercialisation et les recherches. PCL est multiplateforme, il peut être compilé et déployé sous linux, MacOS, Windows et Androide/Ios. C'est un ensemble de bibliothèque C++.

PCL utilise la commande « open NI » pour communiquer avec la Kinect.

Le développement et le financement de PCL :

Le projet est développé par des chercheurs de différentes universités et un grand nombre d'ingénieurs et de scientifiques de différentes organisations.

Le projet est financièrement soutenu par des grands acteurs économiques de différent continent, tels que open perception, Willow Garage, NVidia, Google (GSOC 2011,2012),Toyota, Trimble, UrbanRobotics, Honda Research Institute, Sandia Intelligent Systems et Robotics, Dinast,Optronic, Velodyne, Spectrolab, Fotonic, LeicaGeosystems, Institut national des normes et de la technologie, SouthwestResearch Institute, OcularRobotics, TetraVue,Aptina, Intel et REMAP UCLA.

Pour plus de détails voir :

www.PointClouds.org

*See [http : //www.ros.org/wiki/PCL](http://www.ros.org/wiki/PCL)

And [http : //www.ros.org/wiki/PCL](http://www.ros.org/wiki/PCL)



Résumé :

La classification c'est construire une collection d'objets Similaires au sein d'un même groupe et dissimilaires quand ils appartiennent à des groupes différents. Les algorithmes de classification non supervisées sont souvent utilisés pour étudier des données pour lesquelles peu d'information sont disponible. Il existe une très large famille de méthodes dédiées à la classification non supervisée dont le plus simple est l'algorithme de k-means. Notre objectif de mémoire est d'appliquer une version de k-means sur un base de référence contient 4440 objets 3D. Notre système atteint un taux de bonne reconnaissance de l'ordre 79%.

Mots clés :

Image RGB-D, descripteur 3D, classification, algorithme k-means, kinect.