

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'AUTOMATIQUE

**Mémoire de Fin d'Etudes
De MASTER ACADEMIQUE**
Domaine : Sciences et Technologies
Filière : Génie électrique
Spécialité : **Commande des systèmes**

Présenté par
**Chabane BOURAHLA
Said HAMMADI**

Thème
**Identification automatique
d'objets 3D**

Mémoire soutenu publiquement le ...19.../...07.../ 2016..... Devant le jury composé de :

Mr M.DIAF

Professeur, UMMTO, Président

Mr T.CHELLI

M A (A), UMMTO, Encadreur

Mme Z. HALLICHE

Grade, Lieu d'exercice, Examineur

Mme O.OUAKANDOUR

Grade, Lieu d'exercice, Examineur

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, tout 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 **BOURAHLA** et **HAMMADI**.
- ❖ A nos frères et sœurs.
- ❖ A tous nos amis.

Et enfin nos remerciements à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'étude.

Dédicace

Je dédie ce modeste travail a :

- + Mes parents que j'aime beaucoup.*
- + Mes chers frères et sœurs.*
- + A la mémoire de mes grands parents.*
- + Ma famille, oncles, tantes et leurs enfants.*
- + Mon cher binôme et ami Chabane.*
- + Tous mes amis (es), qui mont beaucoup soutenu .*

HAMMADI Said

Dédicace

Je dédie ce modeste travail à :

- ✚ Mes parents que j'aime beaucoup.*
- ✚ Mon chère frère Abdenour.*
- ✚ Ma chère sœur Lilia.*
- ✚ Ma chère famille.*
- ✚ Ma très chère copine.*
- ✚ Tous mes amis(es).*
- ✚ Tous ceux que je connais.*

CHABANE BOURAHLA

Figure I.1: schéma représente la classification des méthodes de reconstruction 3D.

Figure I.2 : Fonctionnement de la stéréoscopie.

Figure I.3: Principe d'acquisition « Shape From Motion».

Figure I.4 : Différents motifs lumineux en triangulation active : (a) point, (b) ligne, (c) motif.

Figure I.5 : Principe de fonctionnement des caméras Temps-de-Vol.

Figure I.6 : Exemple de caméra à Temps-de-Vol.

Figure I.7 : Forme à partir de texture.

Figure I.8 : Forme à partir de Silhouettes : prise de vues d'un objet suivant deux angles (a) et résultat (b)

Figure I.9: Exemple de reconstruction 3D par « forme à partir de mise au point

Figure I.10 : présentation de la kinect

Figure I .11: composant de la kinect

Figure I.12: les angles de rotation de la kinect

Figure I.13: Le diagramme de fonctionnement de la kinect

Figure I.14 : Exemple de motif infrarouge Projeté par la Kinect. du motif.

Figure I.15: Schéma de la déformation.

Figure I.16 : Calcul de la disparité.

Figure I .17: tableau 1 de comparaison entre kinect pour windows et kinect pour Xbox

Figure I.18: schéma représente les domaines d'utilisation de la kinect

Figure I .19: schéma explicatif représente les domaines d'utilisation de la kinect

Figure I.20 : nuage de point d'un modèle 3D de voiture.

FigureI.21 : image RGB d'une scène réelle (gauche) et l'image recueillie par la caméra 3D Kinect (droite)

FigureII.1 : image RVB (à gauche) et la carte de profondeur correspondante (à droite) capturé avec une caméra **PrimeSense**. Les valeurs de profondeur vont de petites à grandes aussi que les changements de couleur du rouge au bleu.

Figures II.2 : schéma représentant les descripteurs

Figure II.3 : objets échantillonnés à partir de l'ensemble de données RGB-D qui sont triés en fonction de leur taille. De gauche à droite: pomme, tasse de café, bol, casquette et clavier.

Figure II.4 : Top 10 des valeurs propres de matrices de noyau formées par balle et de la plaque.

Figures III.1 : synoptique de la chaîne de reconnaissance de formes

Figure III.2: Images segmentées de notre base de données.

Figure III.3: Masque de chaque image segmentée.

Figure III.4 : un extrait de programme

Figure III.5: des images segmentées dans notre base de données.

Figure III.6 : le mask de chaque image.

Sommaire

Introduction générale.....	1
----------------------------	---

Chapitre I : Acquisition et représentation des images 3D

Partie I : l'état de l'art

1. Introduction	2
2. Triangulation passive.....	3
2.1. La stéréo.....	3
2.2. Forme à partir du mouvement (shape from motion).....	4
3. triangulation active.....	5
4. temps de vol (time of flight).....	6
4.1. Principe de temps de vol.....	6
4.2. Composant d'une camera à temps de vol.....	7
4.3. Avantage de la méthode TOF.....	8
5. autres méthodes.....	9
5.1. Frome à partir de texture (shape from texture).....	9
5.2. Frome à partir de l'ombrage (shape from shading).....	10
5.3. Reconstruction a partir de silhouette (shape from silhouettes).....	11
5.4. Frome à partir de la mise au point (shape from defocus).....	12

Partie II : introduction a la kinect

1. introduction.....	14
2. présentation de la kinect.....	14
3. kinect en chiffre.....	15
4. caractéristiques de la kinect.....	16
5. composant de la kinect.....	17

6. fonctionnement de la kinect.....	18
7. la lumière structurée	21
7.1. Projection de motif connu.....	21
7.2. Analyse de motif déformé.....	21
7.3. Précision de la kinect.....	23
8. comparaison entre la kinect pour windows et kinect pour Xbox.....	24
9. utilisation de la kinect.....	27
10. les conditions matérielles pour utiliser la kinect.....	27
11. points forts et points faibles de la kinect.....	27

Partie III : représentation des images 3D

1. types de données 3D.....	29
1.1. Nuage de point.....	29
1.2. Image de profondeurs.....	30
Conclusion.....	31

Chapitre II : Identification d'objet 3D

1 .introduction.....	32
2. les descripteurs.....	33
2.1. Normale d'un point 3D.....	35
2.2. Descripteur spin.....	35
2.3. Descripteur de taille.....	37
3. Conclusion.....	39

Chapitre III : application des descripteurs RDF

1. Introduction.....	41
2. Chaîne de RDF (reconnaissance de forme).....	41
2.1. Synoptique.....	41
2.2. L'apprentissage.....	42
a. Le prétraitement.....	42
b. La segmentation.....	42
c. Extraction de paramètres.....	44
d. L'apprentissage.....	45
2.3. L'identification.....	45
3. base de données.....	46
3.1. résultat du test.....	46
4. calcul d'indice de similarité.....	48
5. la similarité.....	49
6. conclusion.....	50
Conclusion générale.....	51

Introduction générale

La reconnaissance de forme est une science qui reproduit le plus fidèlement possible la vision humaine, à sa naissance elle utilise des signaux à faible quantité d'information, et elle fait appel à des algorithmes très simples peu performants. Avec les développements des appareils d'acquisition des données et la puissance des ordinateurs actuels, les systèmes de vision par ordinateur peuvent traiter des données plus complexes et riches en information telles que les images 3D acquises par une camera kinect, et développer des programmes puissants et performants.

Notre travail consiste à identifier des objets dans une scène 3D. Ce domaine reste toujours un grand défi pour les chercheurs en vision par ordinateur. Durant les années précédentes, les chercheurs ont développé plusieurs applications de vision 3D, parmi ces applications acquisition de données couleur (RGB) de scène, la profondeur de la scène en temps réel grâce à la kinect et identification automatique des objets 3D dans des nuages de points.

L'objectif de ce projet est de réaliser des applications qui nous permettent d'acquérir des données de l'image 3D en utilisant la technologie Kinect. Qui est un périphérique développé par Microsoft, destiné à la console de jeux vidéo Xbox 360.

Dans ce rapport nous présentons trois parties essentielles. Ainsi, le premier chapitre est consacré à exposer l'état de l'art sur des différentes techniques permettant l'acquisition de données 3D et représenter l'élément nécessaire au projet qui est la Kinect et les outils de développement pour exploiter cette dernière. Le second chapitre est dédié pour l'identification des objets 3D avec l'utilisation de nos descripteurs. Le troisième et dernier chapitre a pour objectif de citer les étapes de l'identification pour faire la reconnaissance 3D. En fin, une conclusion générale est exposée ainsi qu'une bibliographie.

Chapitre 1

Acquisition et représentation des images 3D

Partie I: l'état de l'art**1. Introduction :**

Le monde de la recherche scientifique est en constante évolution. Particulièrement, le domaine de la vision 3D, il évolue de manière exponentiel, notamment grâce à son côté attractif et l'apparition de nouvelles technologies. C'est pourquoi il est primordial de réaliser un état de l'art avant de commencer un projet. En effet, cela peut permettre d'acquérir de nombreuses connaissances dans le domaine mais aussi de réaliser un projet plus cohérent en s'inspirant d'idées déjà réalisées. La lecture d'articles scientifiques peut aussi apporter un moyen de comparer notre travail avec ceux précédemment réalisés.

Pour cet état de l'art, on a retenu plusieurs articles principaux qui nous ont appris beaucoup de choses dans le domaine de la 3D, qui nous ont inspiré pour mener à bien notre projet. Les références de ces articles principaux sont présentées dans la bibliographie.

La figure ci-dessous nous montre la classification des différentes méthodes d'acquisition 3D :

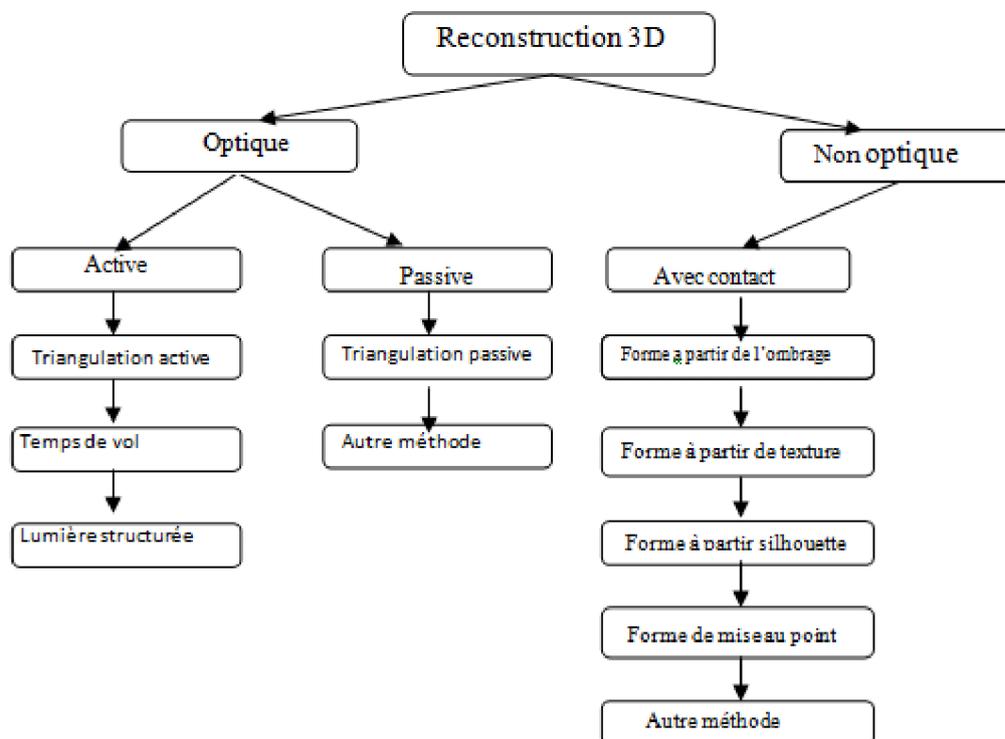


Figure I.1 : schéma représente la classification des méthodes de reconstruction 3D

2. Triangulation passive:

Multiples approches avantageuses utilisent le principe de la triangulation pour l'extraction des informations de profondeur. La triangulation passive ne nécessite aucune source de lumière additionnelle, elle utilise uniquement un ou plusieurs systèmes d'acquisition d'images (caméra, appareil photographique) [1].

2.1. La stéréo :

La stéréo consiste à prendre deux photos au même temps et cela dans deux points de vue différents à l'aide de deux caméras paramétrées (calibrées), étant donné que les deux projections sont sur le même point sur les deux images, la position 3D est retrouvée comme étant l'intersection des deux rayons de projection.

La répétition de ce processus pour plusieurs points donne la forme 3D et la construction des objets de la scène [2].

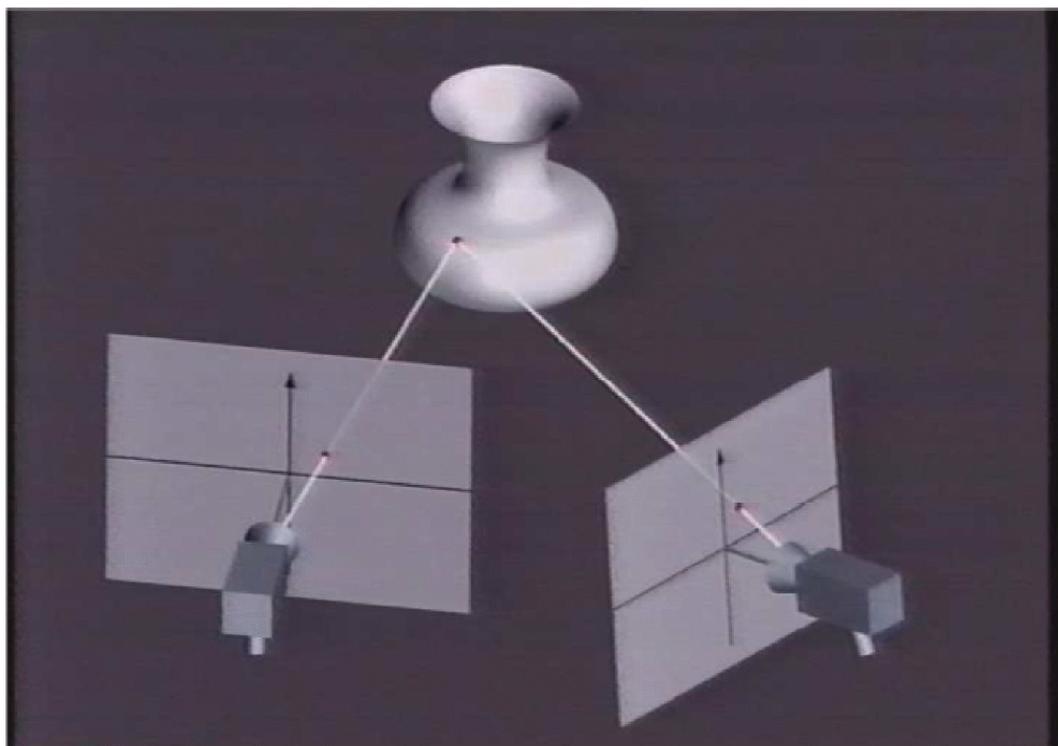


Figure I.2 : *Fonctionnement de la stéréoscopie.*

2.2. Forme à partir de mouvement (Shape From Motion) :

Cette méthode permet de retrouver l'information de surface à partir du mouvement relatif estimé entre l'objet et la caméra. L'acquisition est réalisée soit par une caméra en mouvement réalisant une prise d'images successives ou une vidéo (Figure I.3), 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 de points caractéristiques [3].

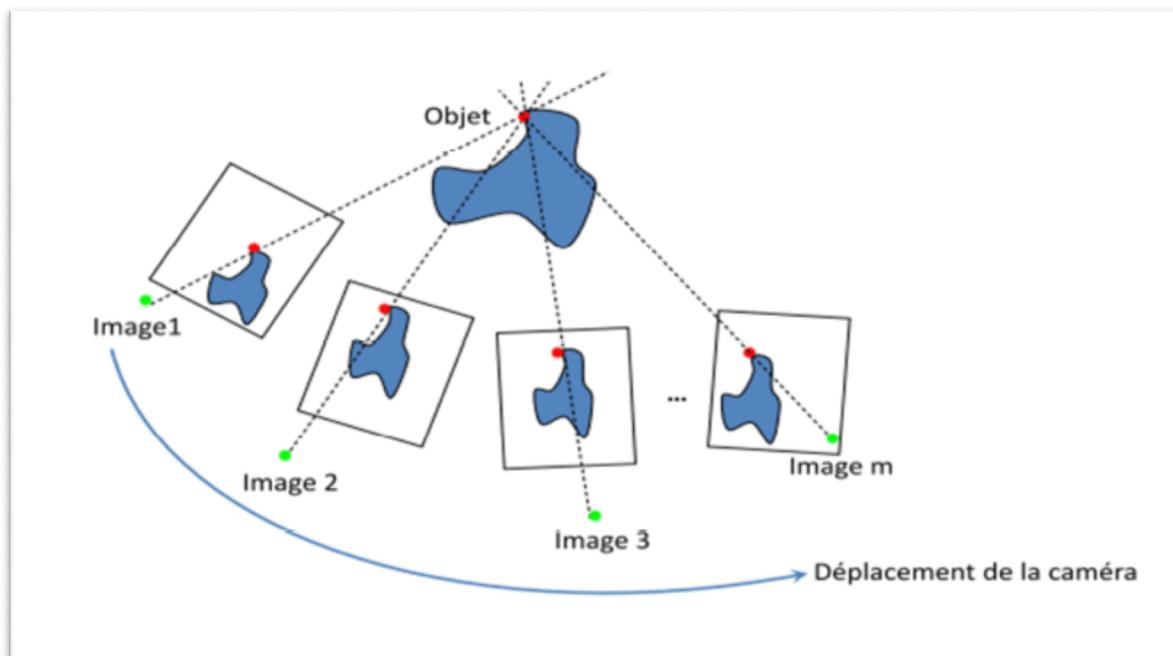


Figure I.3: Principe d'acquisition « Shape From Motion ».

3. Triangulation Active:

La technique de triangulation active se distingue de la technique passive dans la mesure où un éclairage structuré participe à la reconstruction tridimensionnelle de l'objet, cette technique est appelée triangulation car le point laser, l'appareil photo et l'émetteur laser forment un triangle.

Les diverses méthodes de triangulation active peuvent être classées suivant le type du motif lumineux projeté : point, ligne ou motif [1].

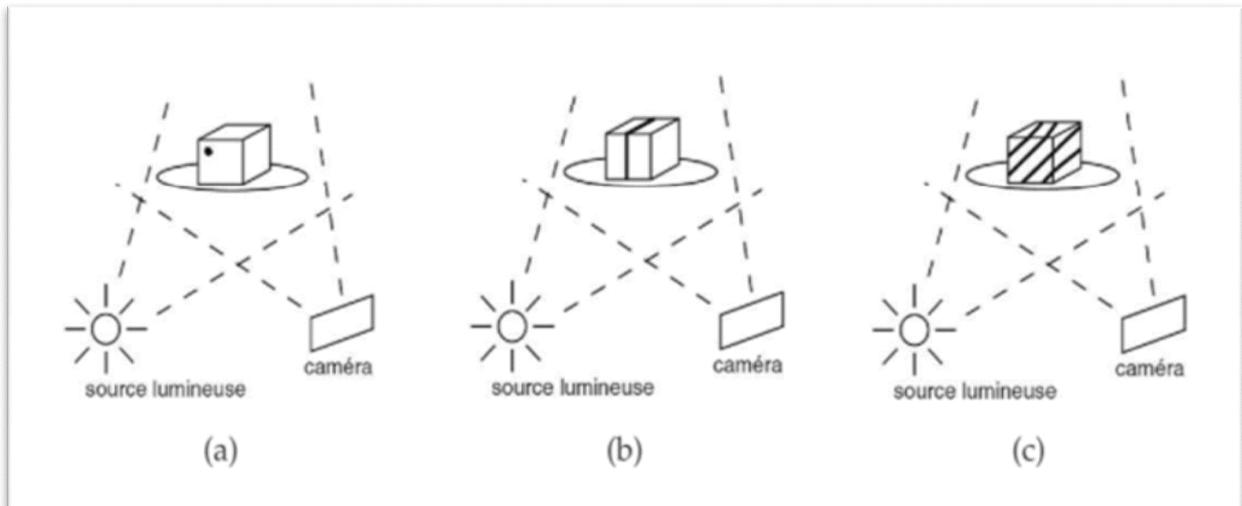


Figure I.4 : Différents motifs lumineux en triangulation active : (a) point, (b) ligne, (c) motif.

4. Temps de vol (Time of flight):

4.1. Principe de temps de vol :

Une caméra fonctionnant sur le principe du temps de vol (anglais : *Time of Flight*,_TOF) permet de mesurer en temps réel le temps de vol d'un signal lumineux entre l'appareil et le sujet pour chaque point de l'image.

La méthode à temps de vol, sont utilisés pour des structures plus grandes pour lesquelles ils sont plus adaptés. Trois données sont nécessaires pour obtenir les coordonnées d'un point, l'angle horizontal, l'angle vertical et la distance entre le scanner et l'objet observé.

Les méthodes les plus courantes pour calculer cette distance sont la détection d'impulsion ou la modulation en amplitude.

La distance D entre le capteur et un point est calculée par la relation suivante :

$$D = v \frac{\Delta t}{2} \quad (1)$$

Où : v est la vitesse du signal.

Δt est le temps mis par le signal pour effectuer l'aller-retour.

$\frac{\Delta t}{2}$ Est Le temps séparant l'objet du scanner.

La précision de ces systèmes est de l'ordre du millimètre sur une distance pouvant aller jusqu'à 300 mètres. Ces systèmes peuvent également être utilisés dans l'industrie, pour le contrôle de pièces de grandes dimensions. Pour ce faire, les caméras TOF illuminent la scène, les objets mesurés par une impulsion lumineuse, et mesurent le temps que cette impulsion prend pour effectuer le trajet entre l'objet et la caméra. Le « temps de vol » de cette impulsion est directement proportionnel à la distance entre la caméra et l'objet mesuré comme le montre **La figure I1**.

Cette mesure de temps de vol est effectuée indépendamment par chaque pixel de la caméra, Permettant ainsi d'obtenir une image complète en 3D de l'objet mesuré [4].

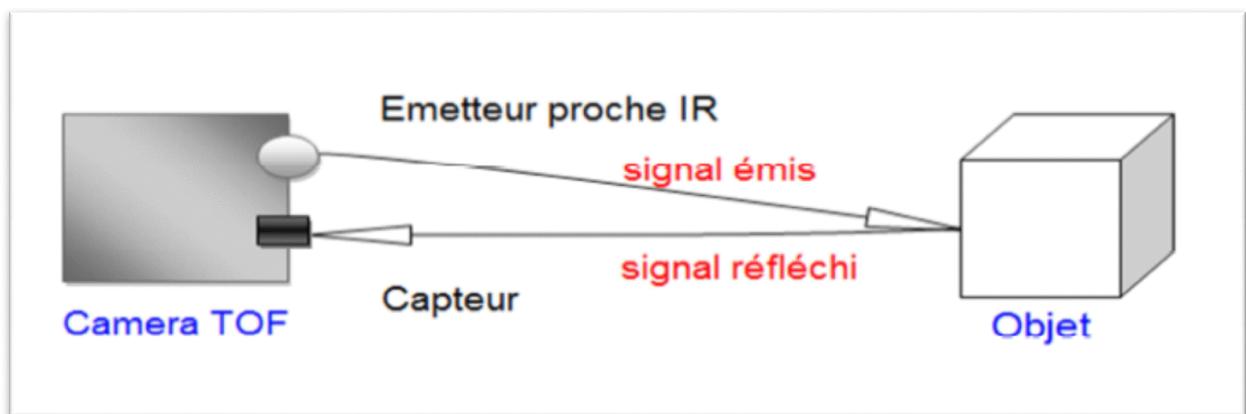


Figure I5 : Principe de fonctionnement des caméras Temps-de-Vol.

4.2. 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.

Objectif: Une lentille rassemble la lumière et des images de l'environnement du réfléchi sur le capteur d'image (focal de la matrice de plan). Un filtre passe-bande optique ne laisse passer la lumière avec la même longueur d'onde que l'unité d'éclairage. Cela permet de supprimer la lumière non pertinente et réduire le bruit.

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 et à l'arrière de la camera.

Le pilote électronique: Les deux unités d'éclairage et le capteur d'image doivent être contrôlés par des signaux à haute fréquences. Ces signaux doivent être très précis pour obtenir une grande résolution.

La figure suivante présente une caméra qui utilise le principe de temps de vol :



Figure I.6 : *Exemple de caméra à Temps-de-Vol.*

4.3. Avantage de la méthode TOF:

Contrairement à la vision stéréo ou systèmes triangulation, l'ensemble du système est très compact: l'éclairage est placé juste à côté de l'objectif, tandis que les autres systèmes ont besoin d'une certaine ligne de base minimum, les caméras à temps de vol aussi sont capables de mesurer les distances dans une scène complète d'un seul coup. Comme les caméras atteignent jusqu'à 160 images par seconde, elles sont idéalement adaptées pour être utilisés dans des applications en temps réel.

Lors de l'utilisation CMOS ou autres détecteurs intégrant ou capteurs qui utilisent la lumière visible ou proche (400 nm - 700 nm), bien que la plupart de l'arrière-plan lumière provenant de l'éclairage artificiel ou la lumière de soleil et supprimée, le pixel doit encore fournir une grande gamme dynamique. La lumière de fond génère également des électrons, qui doivent être stockées.

5. Autre méthodes :

5.1. Forme à partir de texture (Shape from Texture) :

Cette technique permet de reconstruire une forme à partir de sa surface en utilisant les informations contenues dans sa texture. 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 normales à la surface de l'objet est retrouvé à l'aide d'algorithmes complexes valables en théorie mais peu utilisés en pratique.

En effet, la reconstruction tridimensionnelle de la forme n'est possible que si les normales sont assez denses et la surface assez lisse avec une texture régulière.

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 [5].

Dans la **figure I7** :

- a. Image originale
- b. Champ de normales.
- c. Surface reconstruite à partir du champ de normales.
- d. Carte de profondeur

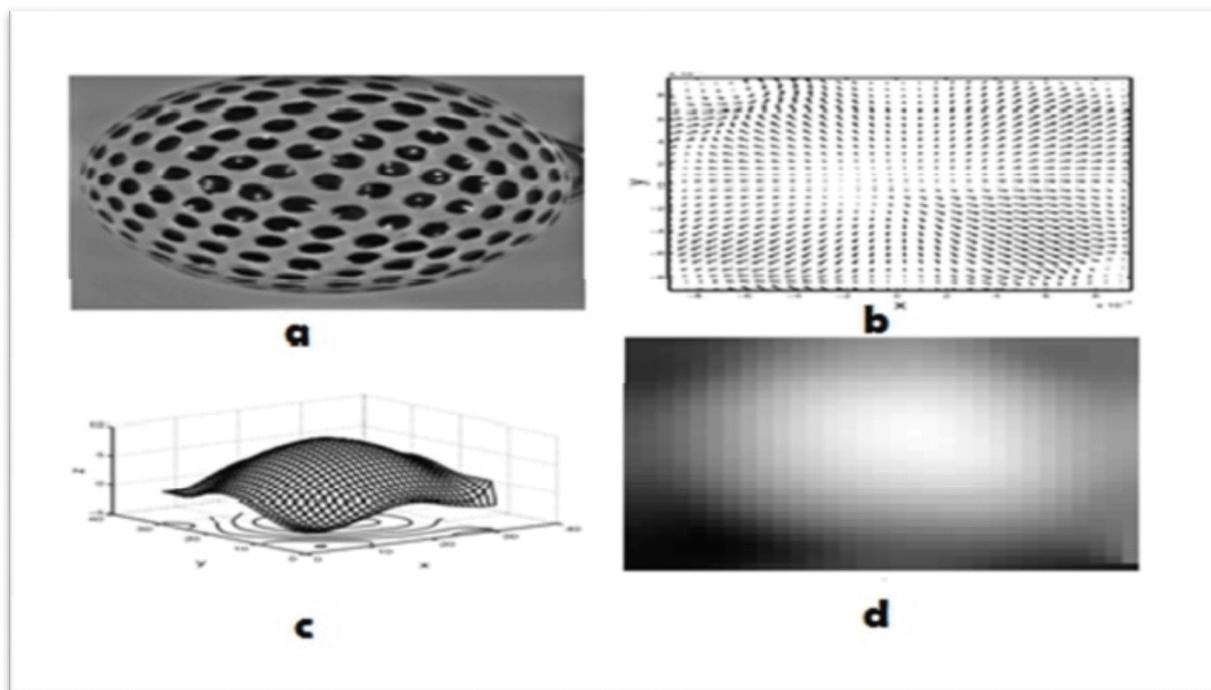


Figure I.7 : *Forme à partir de texture.*

5.2. Forme à partir de l'ombrage (Shape from Shading) :

Les méthodes d'estimation de forme à partir de l'ombrage, emploient une seule caméra et utilisent les variations d'intensité dans l'image pour estimer la forme d'un objet.

L'estimation de la profondeur en chaque point 3D de l'objet est assurée en analysant la brillance de la scène en ce point qui est définie par le niveau de gris de sa projection sur le pixel de plan d'image de la caméra.

La brillance d'un point 3D de la surface de l'objet dépend de l'éclairage de la scène et de la forme de la surface. Elle dépend aussi des propriétés de la réflexion de la surface et de la projection de l'image sur le capteur. Le calcul de la normale en tout point de la surface s'effectue grâce à la minimisation d'une fonction reliant la brillance réelle de la surface de l'objet et la brillance obtenue par estimation de la carte de la réflexion. Des contraintes sur l'intégrabilité et la forme de la surface peuvent être également prises en compte durant le calcul. Cette technique est de faible

coût mais la qualité de la reconstruction n'est pas suffisante surtout en présence de facteurs externes influençant la réflexion de l'objet.

L'intensité de la lumière réfléchiée dépend de son incidence. De plus, les sources lumineuses doivent être suffisamment éloignées de telle sorte que l'illumination soit approximativement uniforme sur toute la surface [6].

5.3. Reconstruction à partir de silhouette (Shape from Silhouettes) :

Cette technologie permet de reconstruire un modèle 3D à partir de plusieurs images d'un objet, prises depuis différents points de vue. Une mire est utilisée afin de calculer la pose de l'appareil photographique pour chaque position.

Chaque image est ensuite segmentée afin de séparer l'objet du fond. La silhouette de l'objet sur chaque image est ainsi obtenue. La combinaison de ces différentes silhouettes permet de générer un modèle 3D texturé.

Pour faciliter les acquisitions, un plateau rotatif peut être utilisé. Cette technique a plusieurs limites. Afin d'obtenir un modèle 3D de bonne qualité, un grand nombre de photographies doit être effectué. Si l'objet présente des surfaces concaves, elles n'apparaîtront pas sur les silhouettes et ne seront donc pas reconstruites. Par contre, elle ne nécessite qu'un appareil photographique (le plateau rotatif évite le déplacement du capteur). Elle donne des résultats satisfaisants pour les applications de visualisation [7].

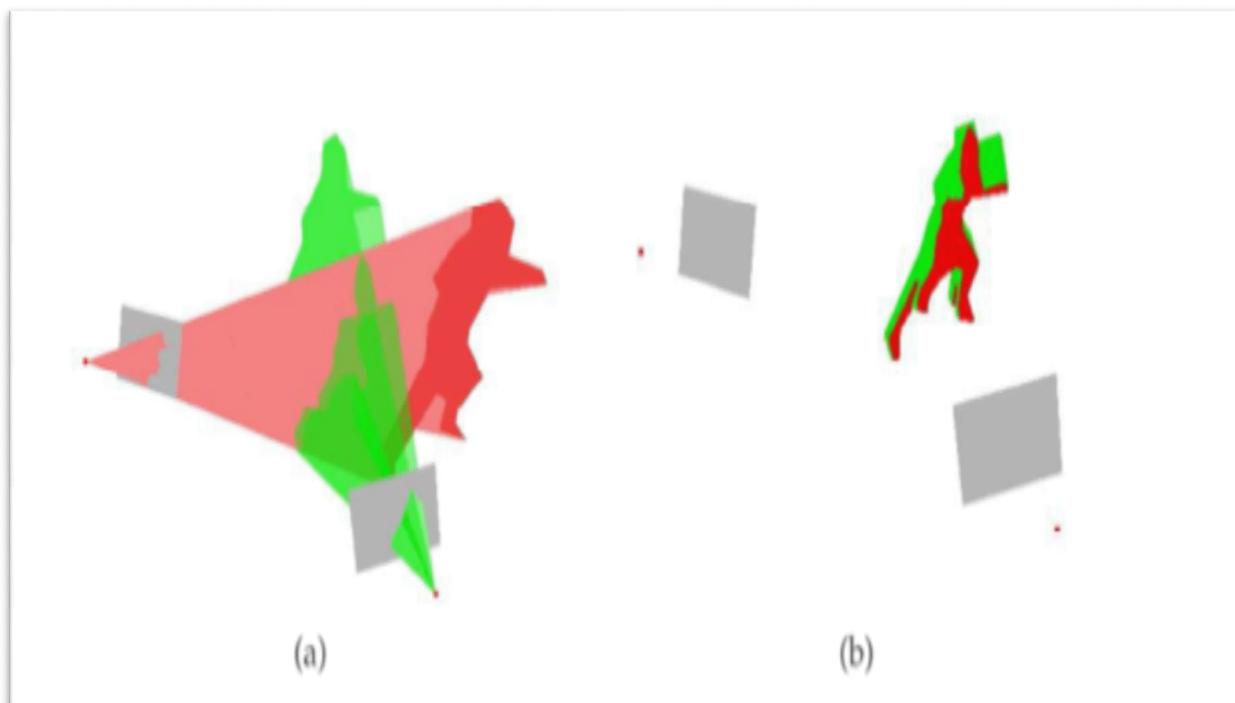


Figure I.8 : *Forme à partir de Silhouettes : prise de vues d'un objet suivant deux angles (a) et résultat (b)*

5.4. Forme à partir de la mise au point (Shape from Defocus) :

Dans le cas de la méthode «Shape from Defocus», on utilise un nombre limité d'images (au minimum deux) acquises avec des paramètres optiques différents afin de ne pas avoir un flux optique identique dans les images.

Pour obtenir l'information de profondeur, cette technique exploite le flux optique perceptible sur les images. Ce flux correspond à un étalement de l'intensité lumineuse correspondant à un point de la scène sur le capteur. Cet étalement est lié à une mauvaise focalisation de l'image sur le plan capteur. C'est l'estimation de cette fonction d'étalement du point qui va permettre de remonter à l'information de profondeur car plus l'étalement est important, plus la distance objet/capteur l'est également.

Le principe est de projeter un motif sur la scène et de comparer ensuite l'image du motif projeté sur la scène à l'image du modèle projeté connue pour déterminer la profondeur de la scène. Elle permet de reconstruire efficacement des surfaces lisses notamment lorsque la radiance

de la scène est contrôlée. Cette méthode est difficilement applicable à des scènes complexes.

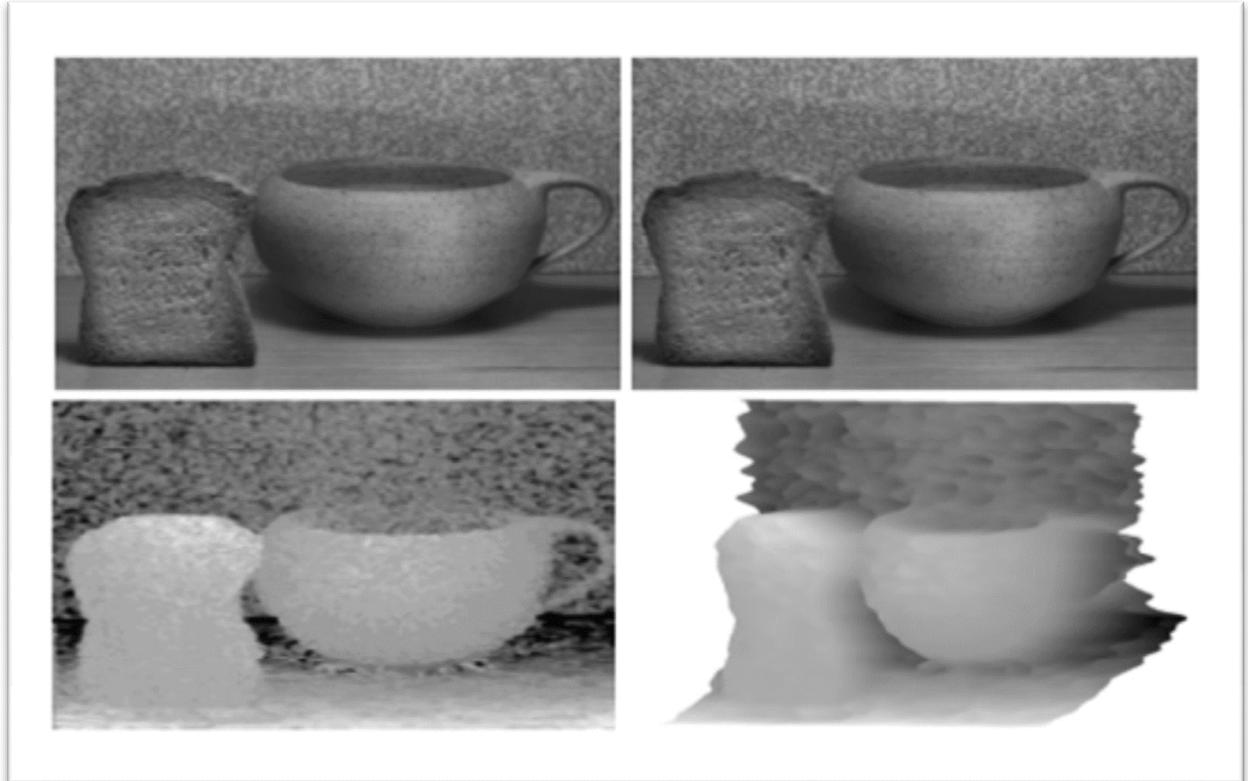


Figure 1.9: Exemple de reconstruction 3D par « forme à partir de mise au point »

Partie II : Introduction a la kinect**1. Introduction :**

Dans cette deuxième partie, on commencera par présenter l'élément principal nécessaire au projet, la Kinect de Microsoft, il s'agit d'un dispositif de détection de mouvement qui était initialement développé pour la console de jeu Xbox 360. Kinect a inauguré une nouvelle révolution dans le monde de jeux, et elle a complètement changé la perception d'un dispositif de jeu. Depuis sa création, elle a continué à briser plusieurs records dans le domaine du matériel de jeux. Pas étonnant Kinect détient le record mondial Guinness pour être le consommateur le plus rapidement vendu

2. Présentation de la Kinect :

Kinect est présenté par Microsoft comme une « expérience de divertissement et de jeu sans contrôleur », il s'agit d'un périphérique destiné à la Xbox 360.

Il permet aux utilisateurs d'interagir avec la console au travers d'une interface simple contrôlable par des gestes ou par des commandes vocales.

Tout cela est rendu possible via un système de suivi et d'analyse du corps humain qui enregistre les mouvements de 48 points du corps en temps réel. Cette identification est permise par le matériel présent sur le périphérique.

Kinect a la capacité unique de « voir » en 3D. Le système Kinect est capable de construire une « carte de profondeur » de la zone en face de la camera. Cette carte est produite entièrement à l'intérieur de la barre de détection, le capteur utilise une technique habile constitué d'un projecteur infrarouge et une caméra qui peut voir les petits points que le projecteur produit.

La figure suivante montre une camera kinect [8] :



Figure I.10 : présentation de la

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, 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.

4. Caractéristiques de la Kinect :**➤ capteur :**

- Caméra couleur et capteur 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 : 0,8 m – 4,0 m.

➤ Flux de données :

- Infrarouge 640x480 à 30 images par seconde.
- RVB 640x480 à 30 images par seconde.
- Audio 16 bits à 16 kHz.

➤ Système de reconnaissance physique :

- Jusqu'à 6 personnes et 2 joueurs actifs (4 joueurs actifs avec le SDK 1.0).
- 20 articulations par squelette.

5. Composant de la Kinect :

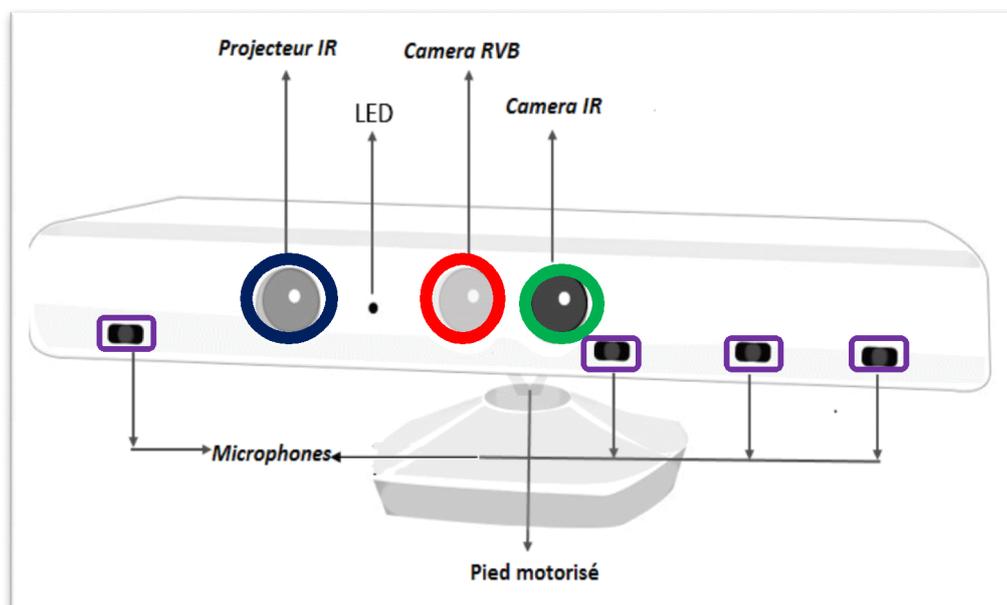


Figure I.11: composant de la kinect

La figure suivante montre une vue de face d'un capteur Kinect. On peut voir les deux caméras au milieu et la source de lumière particulière sur la gauche. Les quatre microphones sont disposés le long de la partie inférieure de la barre de capteur. La Kinect contient également un adaptateur pour l'alimentation externe et un adaptateur USB pour connecter un ordinateur [2].

Kinect se compose des éléments suivants:

=>**Caméra couleur :** La première des deux caméras embarquées dans la technologie Kinect est une caméra couleur RGB « standard » avec un capteur photographique de type CMOS (C'est un composant électronique sensible à la lumière qui va convertir un rayonnement Infrarouge en un signal analogique. Ce signal est ensuite numérisé par un convertisseur analogique-numérique, Afin d'obtenir une image numérique). Pour les férus de détails techniques, elle permet une prise d'image avec une fréquence de 30Hz, en couleur 32bits et en résolution VGA de 640x480 pixels.

=>**Projecteur infrarouge (IR)** : émet en permanence la lumière infrarouge sur les différents points de l'objet. Ces points sont normalement invisible pour nous, mais il est possible de capturer leurs informations de profondeur en utilisant la camera de profondeur IR.

=>**Capteur de profondeur** : La lumière pointillée sur l'objet reflète, et le capteur de profondeur les transforme en informations en mesurant la profondeur qui présente la distance entre le capteur et l'objet.

=> **Moteur d'inclinaison** : Il est utilisé pour changer la position de la caméra et les angles de capteur, pour obtenir la position correcte des objets.

=>**Groupement de microphones**: Le but de micro n'est pas juste de laisser le dispositif Kinect capturer le son mais aussi de localiser la direction de l'onde audio.

=>**LED** : Un voyant est placé entre la caméra et le projecteur infrarouge. Il est utilisé pour indiquer l'état de l'appareil Kinect. La couleur verte de la LED indique que les pilotes de périphériques Kinect sont chargés correctement.

Si vous branchez la Kinect dans un ordinateur, la LED va commencer avec une lumière verte une fois que votre système détecte le périphérique.

6. Fonctionnement de la Kinect :

La gamme visible pour les caméras Kinect est de 43 degrés à la verticale par 57 degrés horizontal. La figure suivante montre une illustration de la plage de visualisation de la Kinect : Le capteur Kinect a la capacité de capturer une vue 3D des objets quelles que soient les conditions d'éclairage de la pièce.

Il utilise un (IR) émetteur infrarouge et un capteur de profondeur infrarouge. La Kinect utilise la technologie Prime Sense pour définir la vision 3D de l'objet.

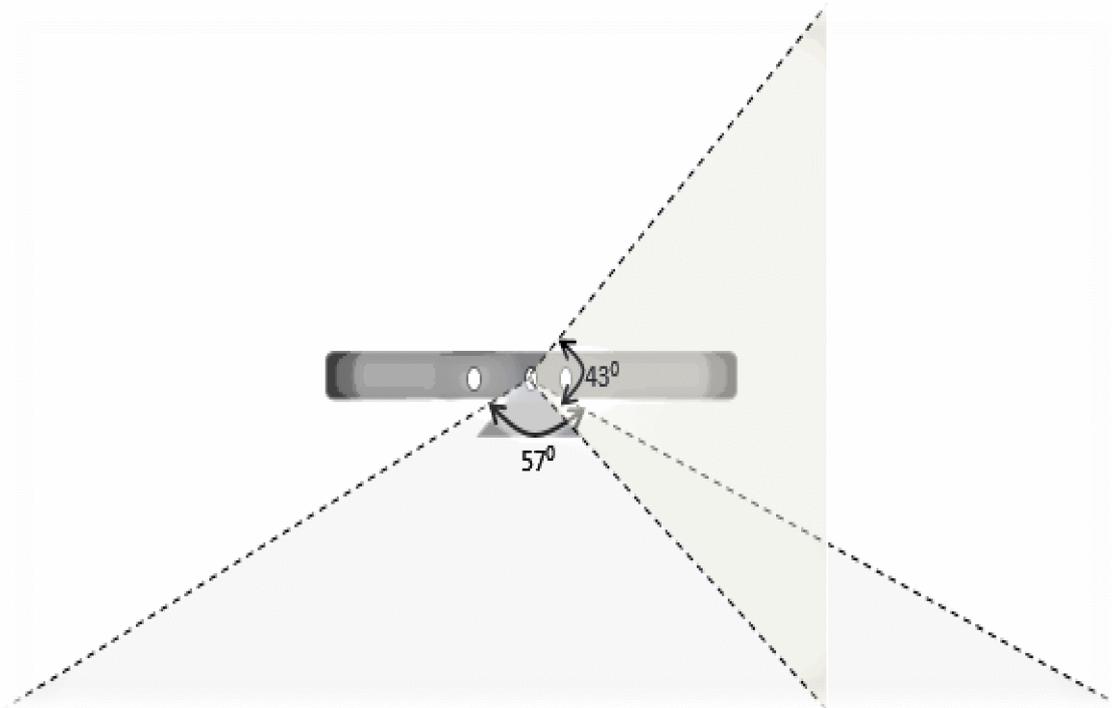


Figure I.12 : Les angles de rotation de la kinect.

Le diagramme ci-dessous montre comment cela fonctionne:

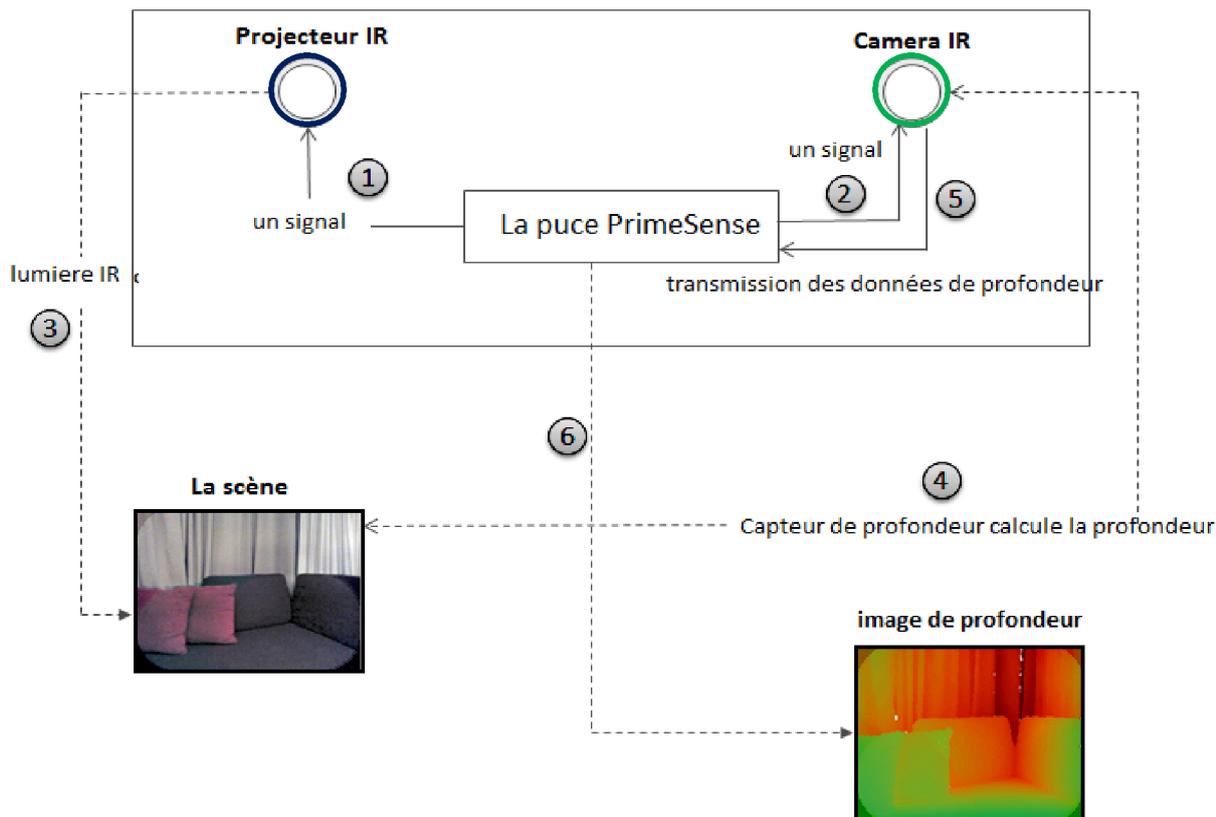


Figure I13: Le diagramme de fonctionnement de la Kinect

La séquence expliquée dans le schéma est la suivante :

Quand il est nécessaire de saisir des données de profondeur, la puce de PrimeSense envoie un signal à l'émetteur infrarouge pour allumer la lampe à infrarouge (1), et envoie un autre signal au capteur de profondeur de l'IR pour lancer la capture de données de profondeur de la scène (2). L'émetteur IR commence l'envoi d'une lumière infrarouge invisible pour les yeux humains (3). Le capteur de profondeur IR commence la lecture des données présumées de l'objet en fonction de la distance de la lumière individuelle des points de réflexion (4), et transmet les données à la puce Prime Sense (5). La puce Prime Sense analyse ensuite les données saisies, et crée une image par image de la profondeur et la transmet comme flux de la profondeur d'une sortie sous forme d'image de profondeur (6).

7. La lumière structurée :

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.

7.1. Projection d'un motif connu :

Tout d'abord, la Kinect projette un motif infrarouge dans la scène. Ce motif étant infrarouge, il est invisible à l'œil nu. Au contact des objets de la scène, le motif se déforme comme le montre la **figure I15** :

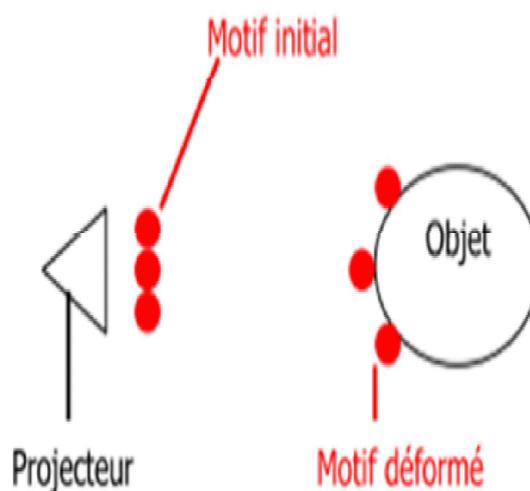
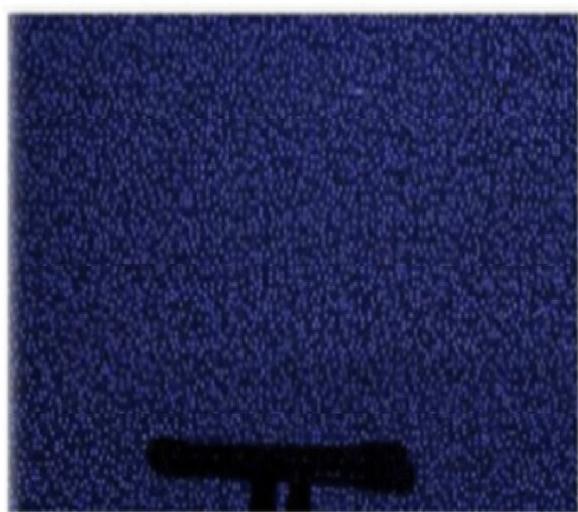


Figure I14 : Exemple de motif infrarouge

Figure I15: Schéma de la déformation

7.2. Analyse du motif déformé :

Le capteur infrarouge étant légèrement décalé du projecteur infrarouge, il peut ainsi capter la déformation du motif pour comparer ce motif avec un motif de référence. Le motif de référence correspond au même motif projeté sur un plan à une distance connue par la Kinect, puis capturé et stocké en mémoire par le capteur. Lorsqu'un point du motif est projeté sur un objet de la scène il est déplacé dans la direction parallèle à l'axe entre la caméra et le capteur infrarouge, appelée ligne de base. Ce

déplacement observé par la caméra est appelé la disparité. Pour chaque point du motif projeté, la disparité est calculée grâce au motif de référence, pour ainsi déduire sa profondeur dans l'espace objet [2].

La figure suivante, schématise le calcul de la disparité d pour un point objet k . Le point du motif projeté correspond au point k et le même point du motif de référence correspond au point o [8].

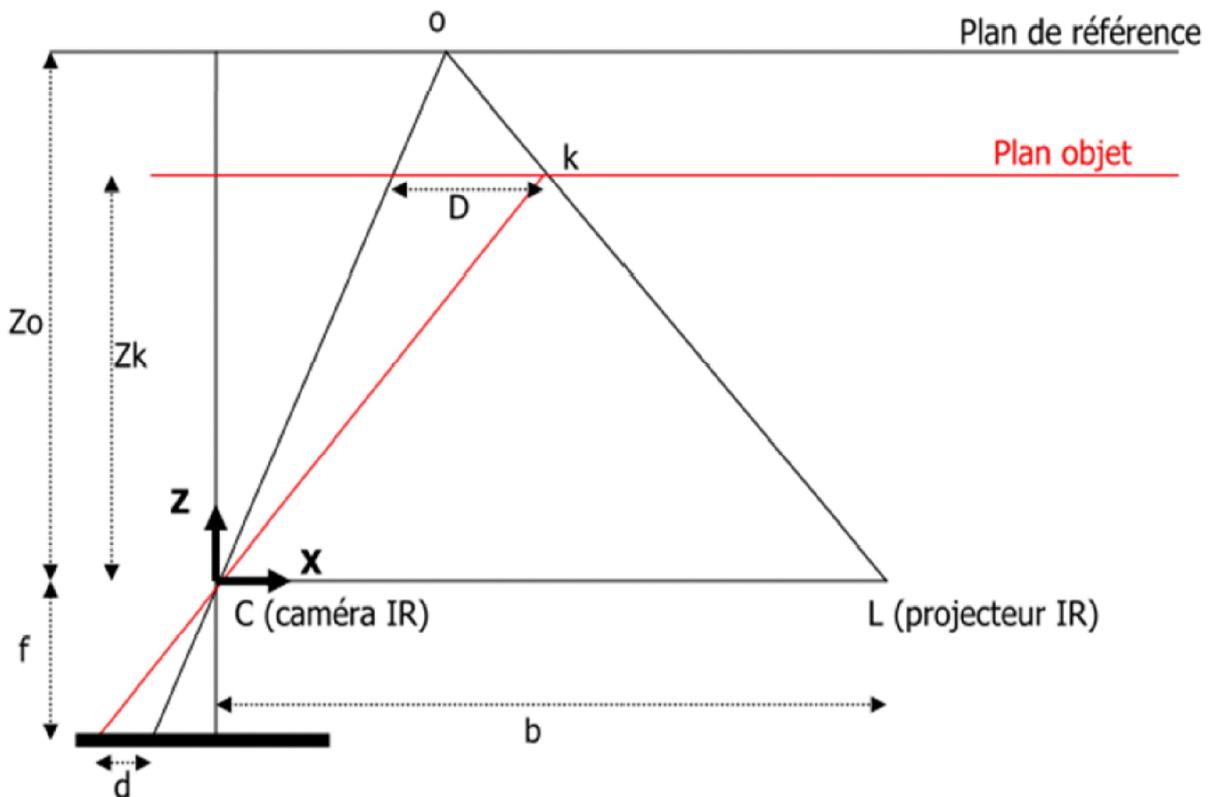


Figure I16 : Calcul de la disparité.

Grâce au théorème des triangles semblables, on peut obtenir les relations suivantes:

$$\frac{D}{b} = \frac{Z_o - Z_k}{Z_o} \quad (2)$$

$$\frac{d}{f} = \frac{D}{Z_k} \quad (3)$$

D : est le déplacement du point du motif projeté ;

b : est la distance entre la caméra infrarouge et le projecteur infrarouge,

Z_o : est la distance du plan de référence,

Z_k : est la distance de l'objet.

f : est la distance focale de la caméra. Z_k , b, et f sont connus par la caméra. En remplaçant D de (2) dans (1), on peut ainsi exprimer la profondeur Z_k .

$$Z_k = \frac{Z_o}{1 + \frac{Z_o}{fb} \times d} \quad (4)$$

7.3. La précision de la Kinect :

Le motif infrarouge est projeté sur les objets de la scène. Plus l'objet est éloigné, plus les points sont diffractés et éloignés les uns des autres. 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.[8]

8. La comparaison Entre Kinect pour Windows et Kinect pour Xbox :

Bien que «Kinect pour Windows» et «Kinect pour Xbox» ont plusieurs différences du point de vue du développeur. Nous devons garder à l'esprit que l'objectif principal de Kinect pour Xbox était d'améliorer l'expérience des joueurs. Développement d'applications n'était pas son objectif principal. En revanche, Kinect pour Windows est essentiellement un dispositif de développement et non pour présenter des jeux. Vous pouvez développer des applications qui utilisent soit la Kinect pour Windows ou la Kinect pour Xbox capteur.

Le capteur Kinect pour Xbox a été construit pour suivre les joueurs qui sont jusqu'à 12 pieds (4,0 mètres) de l'appareille. Mais il ne peut pas suivre les objets que sont très près (80 cm).

Le capteur Kinect pour Windows a un nouveau firmware, qui permet de prendre en charge la suivi d'objets aussi près que 40 cm en face de l'appareil sans perte la précision. En termes de gamme, les deux capteurs se comportent de la même.

Le capteur Kinect pour Windows est utilisé pour applications commerciales, ce qui signifie que si vous développez une application commerciale, vous devez utiliser le Kinect pour appareil Windows pour la production, alors que vous pouvez utiliser Kinect pour Xbox d'apprentissage et de recherche.

9. Utilisation de la Kinect :

Le capteur Kinect pour Windows et le Kinect pour Windows SDK offre des nouvelles occasions pour le développeur de construire un large éventail d'applications.

Ceux-ci peuvent inclure :

- Capture vidéo en temps réel en utilisant le capteur de couleur
- Suivi d'un corps humain et de répondre à ses mouvements et gestes comme une interface utilisateur naturelle
- Mesurer les distances des objets.
- Analyse des données 3D et de faire un modèle 3D.
- Génération d'une carte de profondeur.
- Reconnaître une voix humaine et le développement d'applications mains libres qui peut être contrôlé par la voix, vous pouvez construire un certain nombre d'applications du monde réel qui relèvent d'un autre domaine

Ce qui suit sont quelques exemples, qui vous aideront comprendre l'applicabilité de capteurs Kinect :

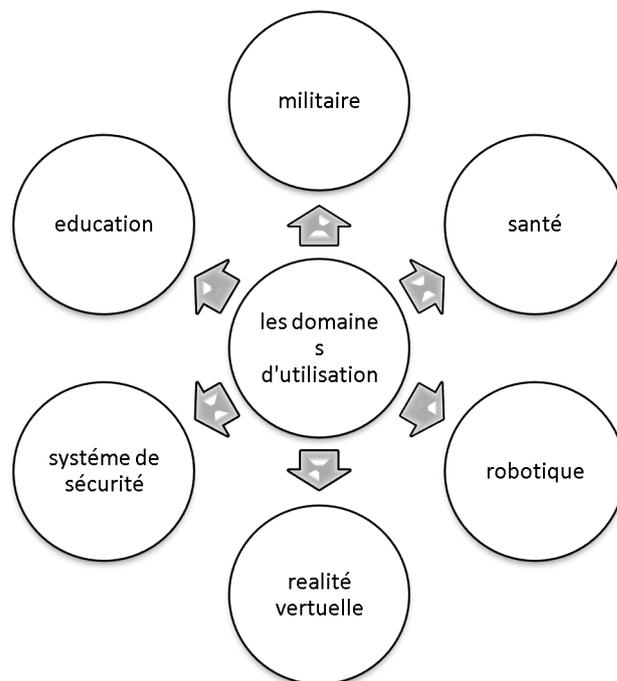


Figure I.18: schéma représente les domaines d'utilisation de la Kinect

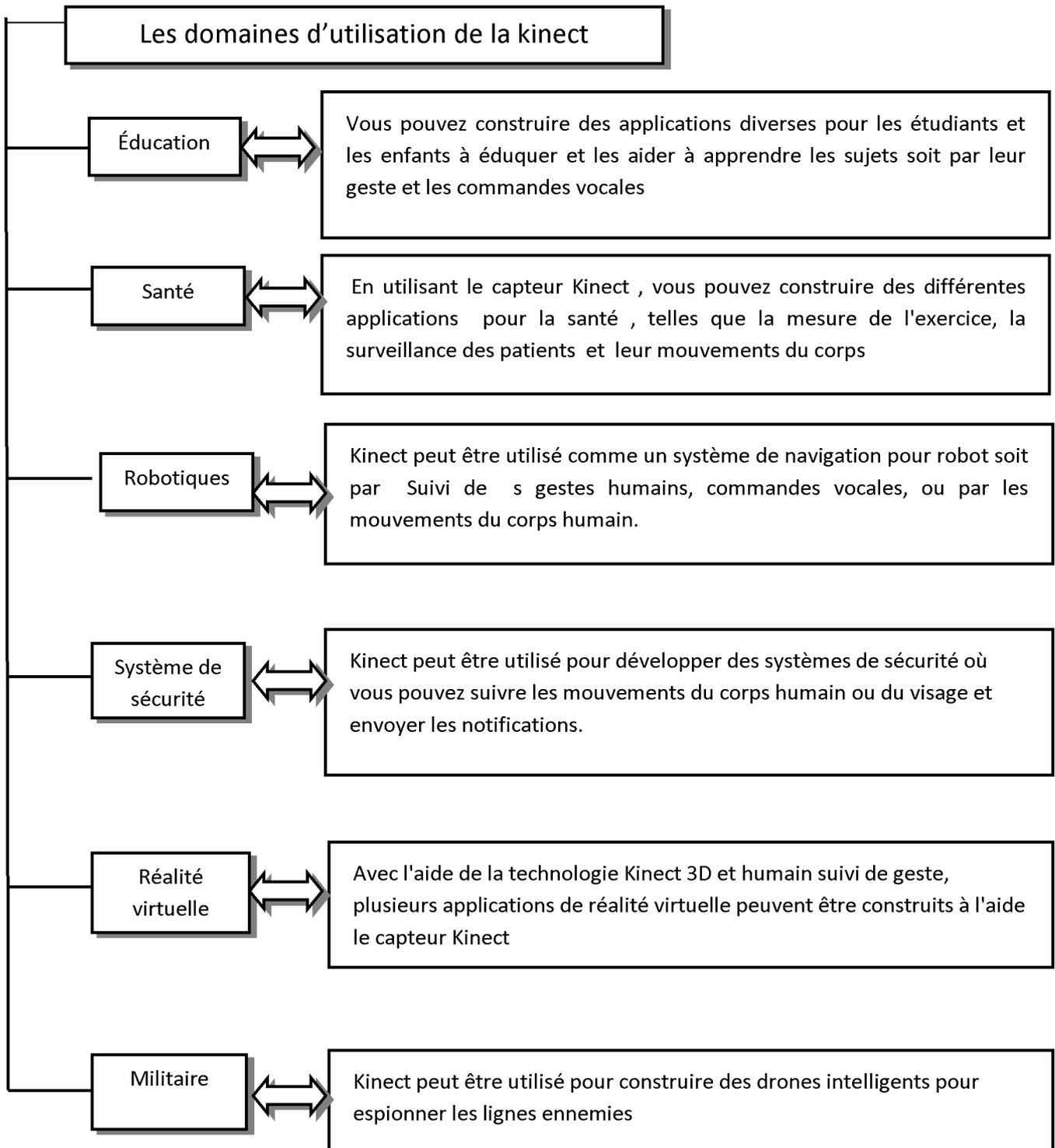


Figure I.19: schéma explicatif représente les domaines d'utilisation de la Kinect

Ce sont quelques exemples des différents domaines où vous pouvez utiliser Kinect, mais à la fin c'est à votre imagination, où et comment vous voulez que ce dispositif va travailler.

10. Les conditions matérielles pour utiliser la Kinect

Elle peut être exécutée sur la plupart du matériel disponible sur le marché. Ce qui suit sont les configurations minimales requises pour le développement avec la Kinect pour Windows:

- A 32 – (x86) ou 64 bits (x64)
- Processeur Dual core (i2) 2,66 GHz ou plus rapide
- USB dédié de bus 2.0
- 2 Go de **RAM**

11. Points forts et points faibles de Kinect :

Points forts et points faibles de Kinect comme tout nouveau concept dans le monde vidéo ludique, Kinect a de nombreux points forts et points faibles. Points forts.

11.1 Les points forts :

- ✓ Kinect innove en amenant une expérience, certes présente sur la Wii mais de manière limitée, à un tout autre niveau en faisant fi de tout contrôleur
- ✓ Kinect ne nécessite aucun calibrage et fonctionne parfaitement avec la majorité des joueurs.
- ✓ Un seul accessoire Kinect est suffisant pour jouer à plusieurs, nul besoin d'acheter un accessoire par joueur ce qui représente une grosse économie.
- ✓ Kinect fonctionne parfaitement dans des conditions peut propices au jeu (dans le noir, avec de nombreux va et viens aux alentours).

- ✓ Kinect proposera des jeux pouvant satisfaire de nombreuses catégories de joueurs, ainsi tous les jeux ne seront pas réservés aux casual gamers.
- ✓ L'utilisation de Kinect sera supportée mais pas obligatoire pour de nombreux jeux, ainsi Kinect ne sera pas imposé aux joueurs préférant le jeu de manière classique

11.2 Les Points faibles :

- ✓ Le prix est quelque peu dissuasif, 150 € ce n'est pas donné.
- ✓ Une crainte subsiste que Kinect devienne au bout du compte lassant, un peu comme le problème rencontré par la Wii actuellement
- ✓ Kinect demande un assez grand espace libre pour pouvoir bouger et gesticuler (le recul conseillé est de 1,80 m pour un joueur et 2,50 m à plusieurs.)
- ✓ Bien que Kinect enregistre les mouvements sans aucune latence, une légère latence (< 1 s) peut être observée. Cela est dû à la manière dont sont programmés les jeux qui attendent qu'un mouvement soit entièrement effectué avant de le restituer.
- ✓ Pour les non polyglottes d'entre nous, la limitation (à priori temporaire) de la reconnaissance vocale à l'anglais, l'espagnol et le japonais peut s'avérer pénalisante.

Partie III : représentation des images 3D**1. Types de données 3D :****1.1. Nuage de points :**

La sortie initiale et commune des scanners laser ou des caméras 3D est un nuage de points (**Figure I20**). ces points sont considérés comme des échantillons de la surface de l'objet et sont arrangés sous formes de système de coordonnées en des triplets (X, Y, Z) avec Z représentant la profondeur. Ce nuage est souvent transformé, pour une meilleure visualisation, en un maillage polygonal avec une reconstruction surfacique. Selon le type du capteur, le nuage de points peut être organisé (structuré) ou désorganisé (non structuré).

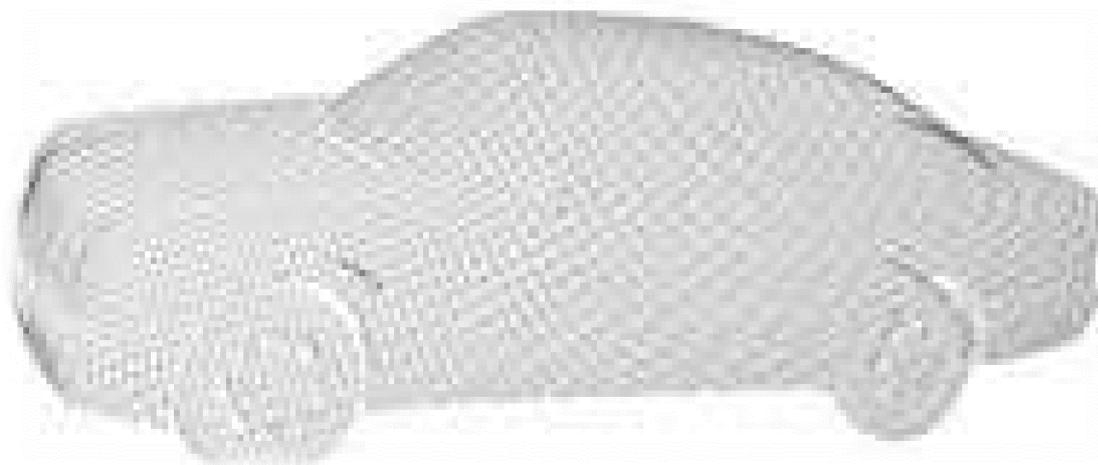


Figure I.20 : nuage de point d'un modèle 3D de voiture

1.2. Image de profondeurs :

Une image de profondeur est une image en niveau de gris dans laquelle l'intensité est inversement proportionnelle à la distance des points de la scène vue sous un seul angle.

Les techniques de mesure de profondeurs (distances) sont multiples .nous citons a titre d'exemple le **Scanner par temps de vol** qui utilise un faisceau laser pour sonder la profondeur. Au cœur de ce type de scanner se trouve un télémètre laser permettant de calculer la distance à la surface de l'objet étudié en mesurant le temps nécessaire au trajet d'aller -retour de l'impulsion du faisceau laser réfléchi. Puisque la vitesse de la lumière est connue, le temps de retour permet de déterminer la distance parcourue par la lumière, qui est deux fois la distance entre le scanner et la surface de l'objet. Évidemment, l'exactitude du scanner par temps de vol dépend de la précision de la mesure du temps de Retour.

La technologie capteur de profondeur 3D (3D depth sensor) à base de **caméra infrarouge** est un autre exemple et elle permet d'obtenir une image représentant les dégagements thermiques émis par l'objet observé. Le fonctionnement est similaire a celui d'une camera RG B sauf qu'uniquement les rayons infrarouges sont émis et filtrés. La scène est bombardée par les rayons infrarouge sont visibles à l'œil.

Une partie de ces rayonnements sera retournée par l'ensemble des objets de la scène. Plus l'objet est loin et plus la quantité de rayonnement infrarouge réfléchi est faible. Ainsi, la caméra infrarouge mesure la distance de l'objet en fonction de l'intensité. Des niveaux de couleurs peuvent être rajoutés en fonction de la température mesurée. La Kinect est une caméra 3D qui comporte une source de lumière infrarouge « structurée », une caméra infrarouge et une caméra couleur (**Figure 121**).

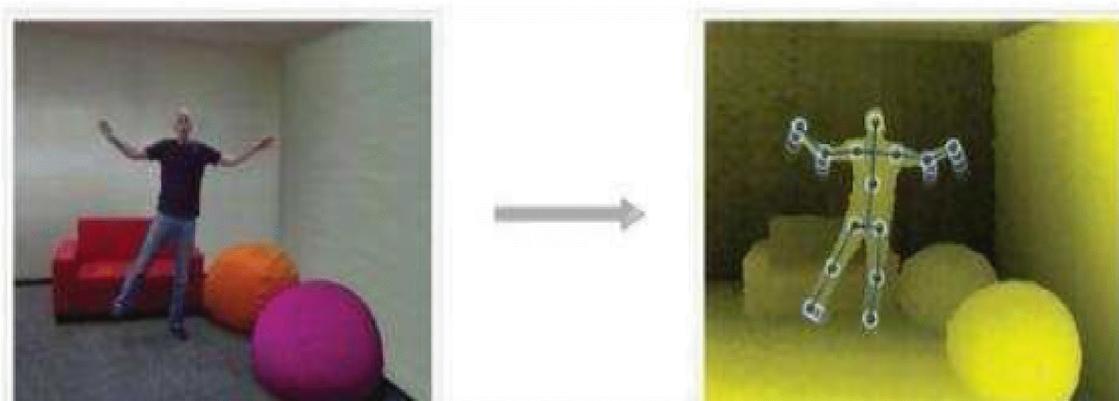


Figure I.21 : *limage RGB d'une scène réelle (gauche) et l'image recueillie par la caméra 3D Kinect (droite)*

Nous mentionnons également la technique utilisant la **stéréoscopie** qui permet, à partir de deux images 2D, de produire une perception du relief. Les systèmes stéréoscopiques utilisent généralement deux caméras vidéo, situées côte à côte et visant la même scène. Un système de caméra stéréo peut être utilisé pour déterminer, avec plus ou moins d'approximation, la profondeur des points dans la scène. Par exemple à partir du point central sur la ligne située entre les deux foyers. Pour mesurer la profondeur à l'aide d'un système stéréo, il est nécessaire, d'abord, de trouver les points qui s'appartiennent dans les différentes images. La résolution de ce problème d'appariement est l'un des principaux problèmes de cette technique (régions d'intensité ou couleur homogène), en plus du coût des capteurs (Multiplié par 2), et surtout de la critique étape de calibration de la paire stéréo [9].

Conclusion :

Dans ce chapitre nous avons présenté un état de l'art sur les méthodes de reconstruction 3D. Chaque méthode possède des avantages et des inconvénients.

Avec le lancement de la Kinect, Microsoft a frappé fort en proposant un condensé de technologie pour un prix particulièrement abordable. Le potentiel de la Kinect ne s'arrête pas à certains jeux et de nombreux projets utilisent déjà la Kinect, Notamment dans le domaine du traitement d'image.

Chapitre 2

Identification d'objet 3D

1. Introduction :

La consommation des caméras de profondeur, tel que Microsoft Kinect, sont capables de fournir des images avec des valeurs de profondeur en temps réel. Une question fondamentale dans l'utilisation des caméras de profondeur est de savoir comment extraire les meilleures caractéristiques des images de profondeur. Motivé par des descripteurs locaux sur les images, en particulier kernel descripteur, nous développons un ensemble de fonctionnalités du noyau sur les images de profondeur tel que la taille du modèle, la forme 3D, et la profondeur des bords dans un cadre unique.

Grâce à des expériences sur la reconnaissance d'objets, nous montrons que :

- Nos caractéristiques locales capture des différents aspects des indices à partir d'une image de profondeur vue qui se complètent les uns les autres;
- Notre noyau se comporte de manière significative (par exemple spin image) ;
- Nous améliorons de manière significative les capacités de la reconnaissance de profondeur et RGB-D (couleur + profondeur).



Figure II.1 : image RVB (à gauche) et la carte de profondeur correspondante (à droite) capturé avec une caméra **PrimeSense**. Les valeurs de profondeur vont de petites à grandes aussi que les changements de couleur du rouge au bleu.

2. Les Descripteurs:

Compte tenu de la grande disponibilité de caméras de profondeur, la question est quelle est la meilleure façon d'extraire des caractéristiques sur une carte de profondeur. Il y a eu beaucoup de travail en robotique sur 3D : Spin Images [10] est un classique exemple de la 3D locale, la Fonction d'Histogramme [11] est un autre exemple d'une fonctionnalité 3D efficace. Ces fonctionnalités 3D, développées sur les nuages de points, ne sont pas toujours bien adaptées aux images de profondeur.

De nombreuses expériences suggèrent que ces nouvelles caractéristiques sont complémentaires et leur combinaison augmente de manière significative, la précision de l'objet de la reconnaissance sur un ensemble de données d'objet RGB-D [12] par rapport aux techniques de l'état de l'art.

Les Descripteurs du noyau ont été initialement proposés pour les images RVB, et n'a pas encore été utilisé pour les cartes de profondeur et des nuages de points 3D. Nous étendons les idées de descripteurs du noyau pour cartes de profondeur et des nuages de points 3D. Nous avons développé deux nouveaux descripteurs profondeur du noyau: le descripteur du noyau de la taille et de spin du noyau descripteur sur les nuages de points 3D. Ces caractéristiques captent des signaux divers mais complémentaires, et leur combinaison améliore la précision de la reconnaissance d'objets de manière significative.

Dans ce travail, nous proposons et étudions une gamme de caractéristiques locales sur une image de profondeur et de montrer que, pour la reconnaissance d'objets ont des caractéristiques supérieures invariants comme Spin Images. Motivé par les derniers développements sur fonctionnalité du noyau sur la base apprentissage [2], nous présentons deux descripteurs de profondeur du noyau qui capture différents indices de reconnaissance, y compris la taille, la forme (Spin Images).

La figure ci-dessous nous montre les différentes méthodes de (kernel descriptor) :

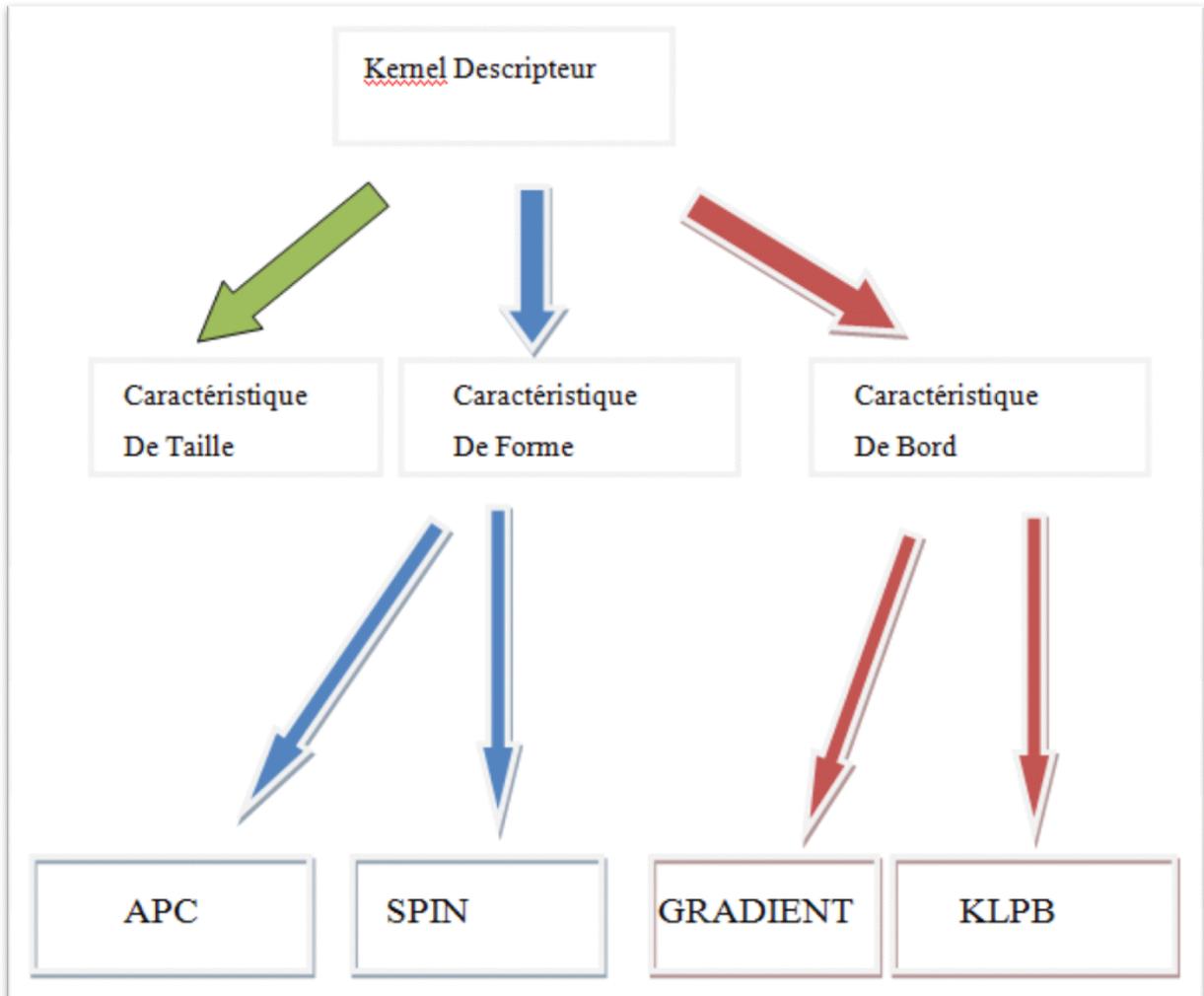


Figure II.2 : schéma représentant les kernel descripteurs.

2.1. Normale d'un point 3 D :

La normale désigne la quantité différentiel d'ordre 0. L'orientation dominante en un point 3D est celle de la normale de la surface local définie par le voisinage du point. L'estimation de la normale en un point 3D constitue une problématique en soit.

On présente ces deux équations pour calculer la normale :

$$\bar{p} = \frac{1}{k} \sum_{i=1}^k p_i \quad (5)$$

Où \bar{p} représente le voisinage d'un point \mathbf{p} :

$$p = \sum_{i=1}^k (p_i - \bar{p})^T (p_i - \bar{p}) \quad (6)$$

L'équation (6) nous donne une matrice 3×3 . On applique l'ACP (Analyse par Composante principale) sur la matrice \mathbf{p} pour trouver la normale \mathbf{n} .

En utilisant la matrice variance covariance à la matrice \mathbf{P} , puis calculer les vecteurs propres. Le min de ces vecteurs est la normale.

2.2 Descripteur Spin :

La forme 3D est un repère stable tant pour la reconnaissance de l'instance et la catégorie. Nous utilisons des caractéristiques de forme pour capturer la forme 3D d'objets. Nous considérons deux caractéristiques de forme sur les nuages de points locaux: caractéristiques du noyau de **IAPC** et le noyau de **spin** descripteur, nous montrons l'intuition pour laquelle le noyau PCA peut fournir des caractéristiques distinctives de forme. Nous

évaluons les matrices du noyau sur les nuages de points 3D d'une balle et une plaque, puis en calcule les valeurs propres de la matrice du noyau résultant, respectivement. Comme nous nous attendons, la distribution des valeurs propres de la balle et de la plaque sont très différents, ce qui suggère que les valeurs propres sont capables de capturer la forme 3D d'objets.

Image Spin [10] est un descripteur de forme 3D locale populaire qui a été largement appliqué à la forme basée sur des problèmes de reconnaissance d'objets. Dans les images de spin, un point dans un nuage de points 3D locale de référence est représenté par le couple (\bar{p}, \bar{n}) formé par sa coordonnée \bar{p} et la surface normale \bar{n} . Les attributs d'image de spin d'un point $p \in P$ représentés par la paire (p, n) sont donnés par le triple $[\eta_p, \varepsilon_p, \beta_p]$

$$\begin{cases} \eta_p = n \cdot (p - \bar{p}) \\ \varepsilon_p = \sqrt{\|p - \bar{p}\|^2 - n_p^2} \\ \beta_p = \arccos(n \cdot \bar{n}) \end{cases} \quad (7)$$

où η_p l'altitude de coordonnées est la signature distance perpendiculaire du point p au plan tangent défini par la paire de la position et la normale (\bar{p}, \bar{n}) .

La coordonnée radiale ε_p est la distance perpendiculaire du point p à la ligne à travers le n normal Et β_p est l'angle entre les normales \bar{n} et n .

Nous avons utilisé ici une variante de l'image de spin qui tient compte de l'angle entre les normales du point de référence et ses points voisins. Ceci a été démontré de meilleures performances que la fonction standard d'image de spin. Nous regroupons les points attributs $[\eta_p, \varepsilon_p, \beta_p]$ dans des caractéristiques de forme locale en définissant le noyau de match suivant :

$$K_{spin}(P, Q) = \sum_{p \in P} \sum_{q \in Q} ka(\bar{\beta}_p, \bar{\beta}_q) k_{spin}([\eta_p, \varepsilon_p], [\eta_q, \varepsilon_q]) \quad (8)$$

Où : $\bar{\beta}_p = [\sin(\beta_p), \cos(\beta_p)]$.

P est l'ensemble des points à proximité autour du point de référence \bar{p} . Noyaux gaussien ka et K_{spin} mesure les similitudes des attributs $[\eta_p, \varepsilon_p, \beta_p]$ respectivement. De la même manière que le descripteur de noyau de taille, nous extrayons le descripteur du noyau de spin de spin match de noyau en projetant le vecteur caractéristique infinie à un ensemble de vecteurs de base finis.

2.3 Descripteur de Taille :

Les caractéristiques de taille capturent la taille physique des objets. L'intuition de la raison pour laquelle la taille physique contribue à la reconnaissance d'objets est illustrée sur la **figureII.03**. Les dimensions physiques des objets sont importants pour la reconnaissance de l'instance depuis une instance d'objet particulier a une taille particulière. Pour la catégorie de reconnaissance, bien que les objets de la même catégorie puissent avoir des tailles différentes, leurs tailles sont généralement limitées à une certaine gamme, qui peut être utile pour la reconnaissance. Par exemple, on peut attendre la caractéristique physique de la taille pour obtenir des performances de reconnaissance parfaite distinction entre les catégories (FigII.03)



Figure II. 03 : objets échantillonnés à partir de l'ensemble de données RGB-D qui sont triés en fonction de leur taille. De gauche à droite: pomme, tasse de café, bol, casquette et clavier.

Pour développer des fonctionnalités de taille, nous convertissons la première profondeur des images à des nuages de points 3D en traçant chaque pixel dans son 3D correspondant de coordonnées vecteur. Pour capturer l'information de taille d'un objet, on calcule la distance entre chaque point et le point du nuage de points de référence (qui est supposé être recadrée autour de l'objet). Plus précisément, Soit P un nuage de points et p le point de ce nuage de référence. Ensuite, l'attribut de distance d'un point $p \in P$ est donnée par :

$$d_p = \|p - \bar{p}\|^2 \quad (9)$$

Pour calculer la similarité entre les attributs de la distance de deux nuages de points P et Q , nous introduisons le noyau de correspondance

$$K_{size}(P, Q) = \sum_{p \in P} \sum_{q \in Q} k_{size}(dp, dq) \quad (10)$$

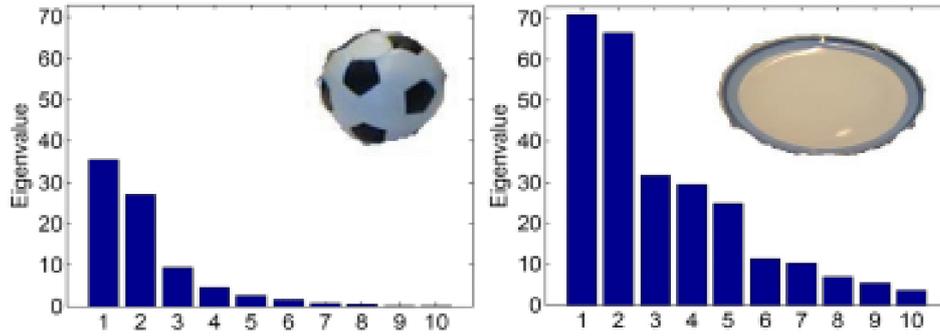


Figure II.04 : Top 10 des valeurs propres de matrices de noyau formées par balle et de la plaque.

Où :

$$k_{size}(dp, dq) = \exp(-\gamma_s \|d_p - d_q\|^2) \text{ avec } (\gamma_s > 0) \quad (11)$$

k_{size} Est une fonction du noyau gaussien.

Comme on peut le voir, le noyau du jeu k_{size} calcule la similitude entre deux ensembles P et Q en additionnant toutes les paires d'attributs à distance. En raison de l'introduction du noyau gaussien, la dimensionnalité du vecteur de fonction sur P est infinie. Adapter les idées de [2], nous projetons cette fonctionnalité vecteur de dimension infinie à un ensemble de vecteurs de base finis, ce qui conduit au descripteur de noyau de dimension finie:

$$F_{size}^e(P) = \sum_{t=1}^{b_s} \alpha_t^e \sum_{p \in P} k_{size}(d_p, u_t) \quad (12)$$

Où u_t sont des vecteurs de base établie de manière uniforme à partir de la zone de support des attributs de distance.

b_s Est le nombre de vecteurs de base.

Et $\{\alpha^e\}_{e=1}^E$ sont les E tops vecteurs propres calculés à partir du noyau analyse en composantes principales.

3. CONCLUSION :

Dans ce chapitre Nous avons présenté deux descripteurs profondeur pour capturer différents indices d'objets dont la taille et la forme, et sa en utilisant les outils mathématiques(PCA) et géométriques (la Normale) ; pour la bonne reconnaissance d'objets et ces descripteurs sont nécessaires pour notre application.

Chapitre 3

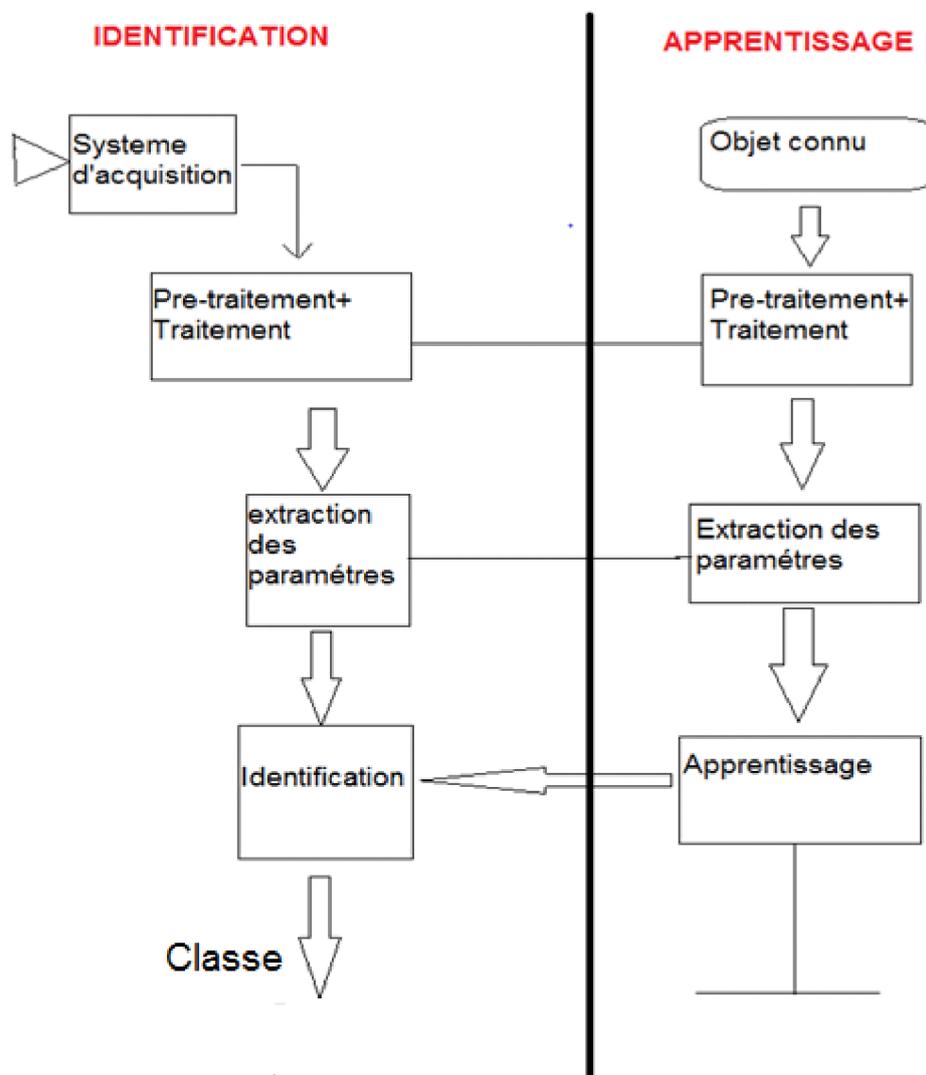
Application des descripteurs RDF

1. Introduction :

Dans ce chapitre nous testons nos descripteurs sur une base de données composé d'image en couleur et en profondeur de plusieurs objets pris dans différents point de vue par notre camera kinect, nous constituerons notre chaine de reconnaissance de forme et comparerons nos résultat.

2 .Chaine de RDF (Reconnaissance De Forme) :

2.1. Synoptique :



figureIII.01 : synoptique de la chaine de reconnaissance de forme.

2.2. L'apprentissage :

L'apprentissage consiste à apprendre la description des objets à partir d'une base d'exemples (galerie) pour construire des classes.

Dans cette partie on introduit 3 étapes pour l'apprentissage :

a. Le prétraitement :

L'étape de prétraitement a pour but de faciliter l'étape de la segmentation en renforçant la ressemblance entre les points appartenant à une même région, ou accentuant la dissemblance entre les points appartenant à des régions différentes. De plus cette élimine les points qui ont des valeurs infinies

b. La segmentation :

La segmentation est l'opération de traitement d'image qui permet de découper l'image en régions d'intérêt sémantique. L'homme est capable naturellement de différencier les objets d'une scène grâce à un mécanisme alliant à la fois vision, compréhension et assimilation. L'élaboration informatique d'un tel algorithme, à un niveau aussi haut, demeure un des défis de la communauté vision. Dans la littérature, les différentes techniques de segmentation peuvent être groupées en quatre classes [13]:

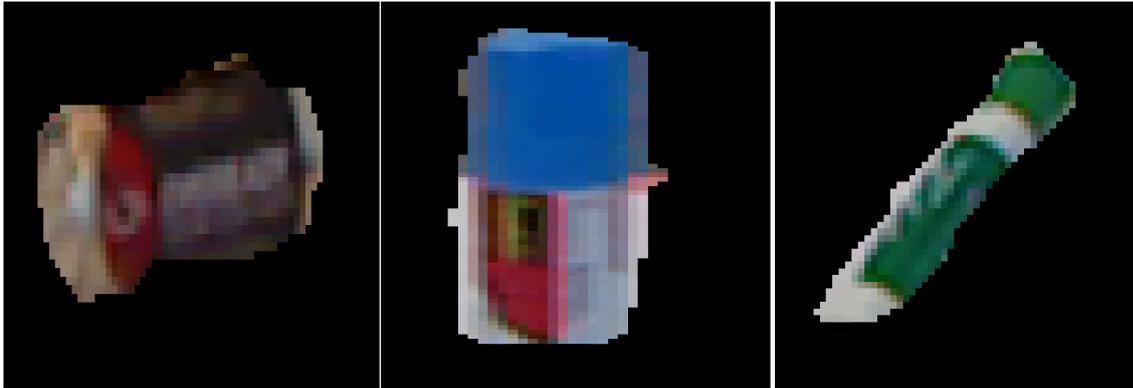
1. Les méthodes basées sur les régions : croissance de région, décomposition/fusion.
2. Segmentation basée sur les contours : segmentation par contours actifs.
3. Segmentation par classification ou seuillage : segmentation par seuillage adaptatif.
4. Segmentation basée sur la composition des trois premières familles.

b.1.1.Segmentation de la région d'intérêt :

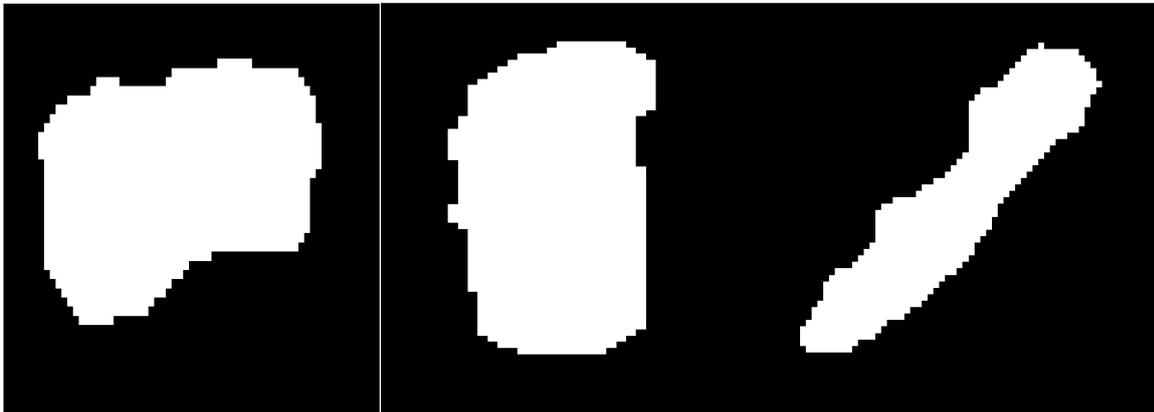
La segmentation d'une scène 3D est aussi un thème très abordé par les chercheurs dans le domaine de la vision par ordinateur. L'utilité de réaliser

une segmentation avant d'analyser une scène est qu'on peut isoler les éléments qui forment la scène.

Voici quelque exemple de la segmentation de notre base de données :



FigureIII. 02: Images segmentées de notre base de données.



FigureIII. 03: Masque de chaque image segmentée.

c. extraction des paramètres :

Dans cette étape on utilise des descripteurs pour extraire les paramètres de l'objet mais dans notre travail on s'intéresse au aux descripteurs de forme et de taille.

c.1 Algorithme du Descripteur :**Lecture des données**

- Lecture des paramètres de la camera
- Lecture de l'image RGB-D

Conversion en point Cloud**Calcule des descripteurs**

Pour $i=1$ jusqu'à la taille du point Cloud.

Pour $j=1$ jusqu'à la taille du point Cloud

$$K_{size}[i, j] = \exp(-\gamma_s \|p_i - p_j\|) \text{ avec } \gamma_s > 0 \quad (13)$$

Calcul de l'indice

Application de l'ACP à K_{size}

```

%calcul des parametres
KerSize = [];
for i = 1:length(impath)
    % convertir en point cloud
    rgbdImg = imread(impath{i});
    l = length(impath{i});
    filetxt = [impath{i}(1:l-14) '_loc.txt'];
    file = fopen(filetxt, 'r');
    loc = fscanf(file, '%d, %d', [1,2]);
    fclose(file);
    pcd = depthcloud(rgbImg, loc);

    [m,n,N] = size(pcd);
    A = reshape(pcd, m*n, N);
    dim = n*m;
    for jj=1:dim
        for ii=1:dim
            KerSize(ii,jj) = exp(-norm(A(ii,:) - A(jj,:)));
        end
    end
    disp(impath{i});
    save(strcat('ker', num2str(i)), 'KerSize');
    clear ('KerSize');
end

```

FigureIII.04: un extrait du programme

d. l'apprentissage :

Cette étape consiste à sauvegarder les classes dans une base de données et chaque classe sera représenté par un indice.

2.3. L'identification :

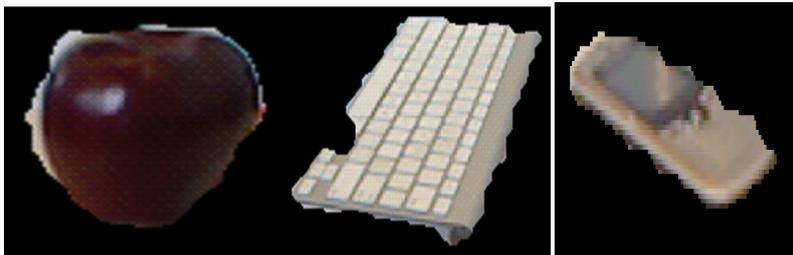
Le système d'acquisition (kinect) prend des images d'objet inconnu qui vont ensuite être traité de la même façon que les objets connus pour le traitement, l'extraction des paramètres et au final ; Ils seront identifiés par le calcul d'indice de similarité.

- Equation d'indice de similarité :

$$K_{size}[i,j] = \exp(-\gamma_s \|p_i - p_j\|) \quad (14)$$

3. base de données :

Notre base de données contient plusieurs classes d'objets (images couleur et profondeur), prises par la kinect et chaque image elle a son image segmenté et son mask et l'image en profondeur et son paramètre. et on va appliquer le descripteur kernel size sur notre base de données afin de calculer les indices et comparer nos résultats.



FigureIII.05 : des images segmentées dans notre base de données.



FigureIII.06 : le mask de chaque image.

3.1. Résultat du Test :

Le tableau ci-dessous montre les résultats trouvés en appliquant le descripteur et l'obtention des indices des objets en utilisant la **ACP** :

Table01 : Pour la classe A(Batterie) :

Batt1	Batt2	Batt3	Batt4	...	Batt19	Batt20
277.96	257.12	302.45	255.45	...	276.75	278.80
0.0066	0.0075	0.0046	0.0069	...	0.0068	0.0057
0.0063	0.0070	0.0046	0.0064	...	0.0065	0.0057
0.0057	0.0063	0.0041	0.0060	...	0.0058	0.0050
0.0055	0.0059	0.0039	0.0055	...	0.0056	0.0045
0.0051	0.0056	0.0038	0.0050	...	0.0052	0.0044
0.0046	0.0053	0.0036	0.0049	...	0.0050	0.0041

Table02 : Pour la classe B(Colle) :

Coll1	Coll2	Coll3	Coll4	...	Coll19	Coll20
346.33	342.66	365.75	357.89	...	337.63	352.65
0.0059	0.0059	0.0058	0.0060	...	0.0058	0.0056
0.0054	0.0054	0.0050	0.0052	...	0.0054	0.0052
0.0051	0.0050	0.0049	0.0051	...	0.0053	0.0051
0.0050	0.0049	0.0047	0.0048	...	0.0048	0.0046
0.0047	0.0047	0.0044	0.0048	...	0.0047	0.0046
0.0045	0.0047	0.0044	0.0044	...	0.0045	0.0045

Table03 : Pour la classe C(Champignon) :

Cha1	Cha2	Cha3	Cha4	...	Cha19	Cha20
393.19	396.29	386.52	405.28	...	399.75	407.44
0.0060	0.0060	0.0057	0.0058	...	0.0049	0.0050
0.0059	0.0057	0.0053	0.0053	...	0.0047	0.0048
0.0058	0.0057	0.0050	0.0050	...	0.0044	0.0044
0.0053	0.0051	0.0046	0.0048	...	0.0043	0.0042
0.0052	0.0051	0.0046	0.0047	...	0.0042	0.0041
0.0049	0.0049	0.0044	0.0046	...	0.0040	0.0039

Table04 : Pour la classe D(Marker) :

Mar1	Mar2	Mar3	Mar4	...	Mar19	Mar20
407.29	360.24	454.75	336.85	...	438.25	462.36
0.0025	0.0031	0.0025	0.0044	...	0.0030	0.0027
0.0025	0.0027	0.0024	0.0039	...	0.0028	0.0025
0.0024	0.0026	0.0022	0.0033	...	0.0025	0.0024
0.0023	0.0026	0.0019	0.0033	...	0.0025	0.0024
0.0022	0.0024	0.0019	0.0029	...	0.0024	0.0022
0.0020	0.0023	0.0017	0.0028	...	0.0024	0.0022

- **Construction de la matrice de confusion et du taux de reconnaissance :**

4. Calcul d'indice de similarité :

a. Distance :

a.1. Distance de Minkowsky :

Cette distance, n'est qu'un cas particulier pour $p = 2$ de la distance de Minkowsky :

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (15)$$

a.2. Distance euclidienne :

La distance la plus connue est la distance Euclidienne, qui définit l'espace cartésien.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (16)$$

(Pour des vecteurs de dimension n). La distance Euclidienne est souvent utilisée au carré.

a.3. Distance de Manhattan :

Pour $p = 1$, on obtient la distance de Manhattan (aussi appelée distance « city-block » ou métrique absolue) :

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (17)$$

a.4. Distance de Chebychev :

Pour $p = \infty$, la distance de Chebychev (aussi appelée distance « Queen-wise » ou encore métrique maximum) :

$$d(x, y) = \max_{i=1}^n |x_i - y_i| \quad (18)$$

5. La similarité :

Une autre méthode de comparer deux vecteurs est de tester leur similarité. Alors que la distance mesure le degré de « différence » entre deux vecteurs, un indice de similarité mesure le degré de « ressemblance » entre deux vecteurs. L'action d'un indice (ou coefficient) de similarité est environ l'inverse de l'action d'une distance : deux vecteurs ayant une grande similarité seront rapprochés (selon la distance) et vice versa.

On va calculer la distance entre chaque objet de chaque classe et en prendra le **min** (des valeurs propres) des distance entre les objets.

Sous MATLAB :

$$dist_{i,j} = norm(V - V_{i,j}) \quad (19)$$

Chaque objet a son indice ou son paramètres principale qui le différencie des autres objets et cette indice qui va être sauvegardé dans notre base de donner et tester avec l'objet inconnu pour faire la reconnaissance, étudier la similarité entre les objets.

Table05 : représentant la matrice de confusion :

	Classe Batterie	Classe colle	Classe champignon	Classe marker	Le taux
Batterie	19	1	0	0	19/20=95%
Champignon	0	14	2	4	14/20=70%
Colle	0	0	14	6	14/20=70%
Marker	1	3	6	10	10/20=50%

Le Taux Total de Reconnaissance est de : 71.25%

Conclusion :

Dans ce chapitre nous avons présenté la chaîne de reconnaissance de forme suivie d'un développement une application de reconnaissance automatique de forme des objets 3D, cette dernière se base sur les paramètres de taille et de forme.

Pour les testes nous avons utilisés une base de données composé de 04 classes de 20 objets chacune (champignon, tube de colle, batterie et marker).

Les résultats obtenus sont de **71.25%** de bonne reconnaissance globale ce qui est assez encourageant, cette application reconnaît très bien les batteries puisque on a un taux de bonne reconnaissance de **95%**, la reconnaissance des markers est faible une chance sur deux pour qu'ils soient bien reconnu. Ces résultats nous permettent de comprendre que ces paramètres de taille et de forme sont largement suffisants pour la reconnaissance certaine objets 3D d'autres non.

Pour l'avenir nous allons introduire d'autre paramètres et de fusion les résultats pour une meilleure perfection.

Conclusion générale

Dans ce mémoire, un état de l'art sur l'acquisition des images 3D a été mise à jour en introduisant les différentes méthodes les plus récentes telle la camera Kinect de Microsoft qui nous a permis d'avoir les données couleurs (RGB) de la scène et la profondeur de la scène par rapport à la camera. Ce système d'acquisition des données 3D nous a facilité à établir une base d'images 3D composée de plusieurs classes d'objets de la vie quotidienne.

Dans le deuxième chapitre nous avons présenté les outils mathématiques (analyse en composante principale) et géométriques (les surfaces normales) ainsi que les paramètres de forme et taille.

Le dernier chapitre de ce travail, présente les étapes que nous avons suivies pour élaborer un système d'identification automatique des objets 3D, en exploitant les paramètres de taille et de forme.

Les tests que nous avons effectués sur une base d'objets 3D constituée de quatre classes (les batteries, champignon, tube de colle, marker) nous ont donné un taux de bonne reconnaissance de 71.25%, ce système peut reconnaître correctement certains objets tels que les batteries (95%) et il est moins efficace pour d'autres tels que les markers (50%), lors des tests qu'on a effectués des problèmes matériels sont soulevés (mémoire ram et microprocesseurs impuissants), ce que ils nous ont fait perdre beaucoup de temps.

Pour améliorer et perfectionner ce système de reconnaissance automatique objets 3D, il faudra agir sur deux fronts, le premier est d'utiliser une station de calcul puissante, et le deuxième est de faire appel à d'autres paramètres locaux pour une précision meilleure ou bien les paramètres globaux pour accélérer le temps d'exécution dans le but d'une application temps réel.

Bibliographie

- [1]** Thèse de B.Loriot ; (Automatisation de l'Acquisition et des Post-traitements en Numérisation 3D) page 6.
- [2]** Thèse de T.Moons, M. Vergauwen, et L. Van Gool ; (3D reconstruction from multiple images) page 8.
- [3]** Thèse de B. BILLIOT ; (Conception d'un dispositif d'acquisition d'images agronomiques 3D en extérieur et développement des traitements associés pour la détection et la reconnaissance de plantes et de maladies) page 23.
- [4]** Thèse de A. Belhedi ;(Modélisation du bruit et étalonnage de la mesure de profondeur des caméras Temps-de-Vol) page 5-22.
- [5]** Un livre de D. Nitzan ;(Three dimensional vision structure for robot applications).
- [6]** Thèse de B.Horn et M.J. Brooks ;(Shape from Shading).
- [7]** Thèse de Y. Xiong and S. Shafer ;(Computer Vision and Pattern Recognition) .
- [8]** Thèse de M. DEVANNE ; (Modélisation 3D d'un corps humain à partir de caméras 3D Kinect) page 18.
- [9]** Thèse A. Shaiek ;(de Reconnaissance d'objets 3D par points d'interet)page27-29
- [10]** Thèse de A. Johnson and M. Hebert ; Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5), 1999.
- [11]** Thèse de R. Bogdan Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In IEEE International Conference on Robotics and Automation, 2009.

[12] Thèse de K. Lai, L. Bo, X. Ren, and D. Fox. A Large-Scale Hierarchical Multi-View RGB-D Object Dataset. In IEEE International Conference on Robotics and Automation, 2011.

[13] Thèse de Nizar SALLEM, Michel DEVY,(Modélisation d'Objets 3D en vue de leur reconnaissance et leur manipulation par un robot personnel) page 3

RESUME

La reconnaissance de forme est une science qui reproduit le plus fidèlement possible la vision humaine, à sa naissance elle utilise des signaux à faible quantité d'information, et elle fait appel à des algorithmes très simples peu performants. Avec les développements des appareils d'acquisition des données et la puissance des ordinateurs actuels, les systèmes de vision par ordinateur peuvent traiter des données plus complexes et riches en information telles que les images 3D acquises par une camera kinect, et développer des programmes puissants et performants.

Notre travail consiste à identifier des objets dans une scène 3D. Ce domaine reste toujours un grand défi pour les chercheurs en vision par ordinateur. Durant les années précédentes, les chercheurs ont développé plusieurs applications de vision 3D, parmi ces applications acquisition de données couleur (RGB) de scène, la profondeur de la scène en temps réel grâce à la kinect et identification automatique des objets 3D dans des nuages de points.

L'objectif de ce projet est de réaliser des applications qui nous permettent d'acquérir des données de l'image 3D en utilisant la technologie Kinect. Qui est un périphérique développé par Microsoft, destiné à la console de jeux vidéo Xbox 360.

Dans ce rapport nous présentons trois parties essentielles. Ainsi, le premier chapitre est consacré à exposer l'état de l'art sur des différentes techniques permettant l'acquisition de données 3D et représenter l'élément nécessaire au projet qui est la Kinect et les outils de développement pour exploiter cette dernière. Le second chapitre est dédié pour l'identification des objets 3D avec l'utilisation de nos descripteurs. Le troisième et dernier chapitre a pour objectif de citer les étapes de l'identification pour faire la reconnaissance 3D.

Les testes que nous avons effectués sur une base d'objets 3D constituée de quatre classes (les batteries, champignon, tube de colle, marker) nous ont donné un taux de bonne reconnaissance de 71.25%, ce système peut reconnaître correctement certaines objets tels que les batteries

(95%) et il est moins efficace pour d'autres tels que les markers (50%), lors des tests qu'on a effectués des problèmes de matériels sont soulevés (mémoire ram et microprocesseurs impuissants), ce que ils nous ont fait perdre beaucoup de temps.

Pour améliorer et perfectionner ce système de reconnaissance automatique objets 3D, il faudra agir sur deux fronts, le premier est d'utiliser une station de calcul puissante, et le deuxième est de faire appel à d'autres paramètres locaux pour une précision meilleure ou bien les paramètres globaux pour accélérer le temps d'exécution dans le but d'une application temps réel.

Mots clefs

Identification automatique d'objets 3D.

Reconstruction d'objets 3D.

Traitement d'image 3D.

Camera temps de vol.

Reconnaissance avec Kinect.

Descripteur Taille et Forme.

Descripteur Spin images.

Etat de l'art sur la 3D.

Synoptique de la reconnaissance de forme 3D.