الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DES SCIENCES ET DE LA TECHNOLOGIE

Département du Génie Électrique et de l'Informatique



Projet de Master

Présenté pour l'obtention du Diplôme
De Master en Génie Électrique

Spécialité : Systèmes de Télécommunication

THEME

SEGMENTATION DES IMAGES RGB-D

Proposé et dirigé par,

Mr. TCHOKETCH-KEBIR Sofian

Préparé et présenté par,

Mr. BENSALEM Nabil

Mr. SKENDER Mohamed Amine

Dédicas

Je dédie se modeste travail à mes chères parents qui m'ont beaucoup encouragé et soutenu pendant tout leur vie pour voir ce jour-là et pour ces précieux conseils ainsi pour ces recommandations

*A l'Etoile de ma Vie mon petit frère '**Mohamed Takiedine**'*

*A Mon frères '**Riad**'*

*A mes sœurs '**Amira et chourouk**'.*

A tout ma famille, grande et petite sans exception.

*A mes amis **Anis, Fayçal, Zouhir, Hassen, Mohamed, B.Lyes, Ali, Youcef Hichem et Redouane** .*

Et Mon** camarade **AMINE

BENSALEM NABIL

Dédicas

Je dédie se modeste travail à mes chères parents qui m'ont beaucoup encouragé et soutenu pendant tout leur vie pour voir ce jour-là et pour ces précieux conseils ainsi pour ces recommandations,

*A mes très chères frères '**Belkacem , Hcene et Hocine.***

*A mes tantes et mes oncles spécialement **Omar** et sa petite famille.*

*A mes cousins **Ali khodja Merouane, Skender Salah Eddine** et **Abderezak.***

A tout ma famille, grande et petite sans exception.

*A mes amis **Abdelrahmane, Ahmed, Mohamed, Abdelhak, Amine, B.Lyes, Lyes, Hichem.***

Et surtout mon amie d'enfance Mohamed Redouane BEN BAHA

*A **Mlle. ZAOUIDI Khadîdja.***

A Mlle. H.BENTERKIA.

Et Mon** camarade **Nabil

SKENDER MOHAMED AMINE

Remerciement

Nous tenons tous d'abord nos sincères remerciements à notre créateur **DIEU** le tout puissant, nous le louons et le remercions pour tout ce qui a donné et qui nous a incité à acquérir le savoir et le réussir. C'est à lui que s'adresse toute notre gratitude

À travers ce modeste travail, nous tenons à remercier vivement notre promoteur

« Mr TCHOKETCH-KEBIR Sofiane» pour l'intéressante documentation qu'il a mis à notre disposition, pour ses conseils précieux et pour toutes les commodités et aisances qu'il nous a apporté durant notre étude ainsi qu'au cours de la réalisation de ce projet de fin d'étude.

Nos remerciements les plus vifs s'adressent aussi à messieurs : le président et les membres de jury d'avoir accepté d'examiner et d'évaluer notre travail.

Nous exprimons également notre gratitude à tous les professeurs et enseignants qui ont collaboré à notre formation depuis notre premier cycle d'étude jusqu'à la fin du cycle universitaire.

Sans omettre bien sûr de remercier profondément tous ceux qui ont contribué de près ou de loin à la réalisation du présent travail

Résume

ملخص: مع تقدم العالم في مجال الملتيميديا والتكنولوجيات الحديثة كان من الضروري مواكبة التحديثات اللازمة في هذا العصر.

نناقش في هذا العمل دراسة تقنية تقسيم الصورة الثنائية و الثلاثية الأبعاد و إعادة تركيبها مع معرفة حدودها و أبعادها كذا تحديد تموقع الأجسام بدقة. و هذا بالاعتماد في الاساس على تقنية تجزئة الرسم البياني وأيضا بعض الخوارزميات التي تمكننا فيما بعد من إعادة الهيكلة أو البناء.
لنستطيع في الأخير تطبيق عملنا هذا في جميع مجالات الصورة الرقمية ومجال الروبوتيك خاصة.

كلمات مفتاحيه : صورة ثنائية، صورة ثلاثية الأبعاد، ملامح، تجزئة الحدود، أبعاد، الرسم البياني، الروبوتيك.

Résumé: Avec l'avancement du monde de Multimédia et la technologie moderne Il était nécessaire d'exploiter les mises à jour fondamentales qui sont en ligne avec cet âge.

Nous abordons dans ce travail pour étudier la technique de segmentation pour l'image simple et tridimensionnel et à la reconstruction et connaitre ces bords ces dimension, Ceci est basé principalement sur la méthode de la **segmentation par histogramme** puis des algorithmes pour la reconstruction 3D.

A la fin en peux appliquer notre travail dans le domaine de l'imagerie Numérique et celui des robots.

Mots clés : Bords, Dimension, Histogramme, Image 2D et 3D, Reconstruction, Robots, Segmentation.

Abstract : With the advancement of technology in the field of media and modern technology was needed to operate the necessary updates that are in line with this age .

We discuss in this work to study the segmentation technique for Simple and three-dimensional image and restructuringand knowing these edges and the dimension, this is based primarily on the method of **segmentation by histogram** and simple algorithms for 3D reconstruction

At the end can apply in our work in the field of robots

Keywords: Image 2D and 3D, Segmentation, Histogram, reconstruction, size, edges, Robots

Liste des figures

Figure I.1 : *Représentation d'une image numérique.*.....4

Figure I.2 : *Les types de pavages*.....5

Figure I.3 : *Les types de Maillage*.....5

Figure I.4 : *Les types de connexité*.....6

Figure I.5 : *Résolution spatiale*.....7

Figure I.6 : *Résolution tonale*.....7

Figure I.7 :*Représentation d'histogramme*.....8

Figure I.8 :*Echantillonnage d'image* 10

Figure I.9 : *Exemple image matricielle*.....13

Figure I.10 : *Example image vectoriel*.....13

Figure I.11 : *Exemple d'image en couleur*.....14

Figure I.12 : *Exemple d'image binaire*.....14

Figure I.13 : *Exemple d'image en niveau de gris*.....15

Figure I.14 :*Schématisation d'une matrice de voxels en nuances de gris*.....16

Figure II.1 : *Chaine de traitement d'une image couleur*.....20

Figure II.2 : *Stratégie scalaire*.....21

Figure II.3 : *Stratégie marginale*.....21

Figure II.4 :*Stratégie vectorielle*.....22

Figure II.5 :*Segmentation de l'image (a) par la méthode division de régions en donnant l'image segmentée (b)* 25

Figure II.6 :*Image original (a) et son GAR (b)* 26

Figure II.7 :*Structure pyramidal du quadtree.* 27

Figure II.8 :*Exemple de diagramme de de Voronoï*.....28

Figure II.9 :*Nuages de points des pixels de l'image originale (a) dans l'espace utilisé (b)*.....29

Figure III.1 :*Schéma fonctionnel d'un système informatique pour la vision artificielle*.....35

Figure III.2 :*Une reconstruction projective à droite*.....36

Figure III.3 : *Une reconstruction affine à droite*.....37

Figure III.4 :*Une reconstruction euclidienne à droite*.....37

Figure III.5 :*Synoptique général de la vision stéréoscopique*.....39

Figure III.6 :*Une configuration possible d'un cube triangule*.....40

Figure III.7: *La configuration miroir d'un cube triangule*.....40

Figure III.8 : *Méthode de reconstruction*.....42

Figure III.9 : *Application de la SFS sur une image 2D*.....43

Figure III.10 : *Polygones triangulaires d'un lapin et d'une poire*.....44

Figure III.11 : *Squelettes topologiques de quelques objets*.....44

Figure III.12: *Processus de reconstruction 3D à partir d'un mouvement*.....47

Figure III.13: *Méthode par space de l'objet*.....47

Figure IV.1: Un petit récapitulatif des technologies embarquées dans Kinect.....51

Figure IV.2: : Représentation du spectre électromagnétique.....53

Figure IV.3: Contrairement aux couleurs primaires en peinture (cyan, magenta et jaune).....53

Figure IV.4: Capteur photographique de type CMOS développé par Sony.....54

Figure IV.5: Fonctionnement d'un capteur CMOS pour les appareils photos numériques 55

Figure IV.6: Capteur photographique de taille réduite pour les téléphones mobiles 55

Figure IV.7: Illustration des émissions infrarouges non visible par notre œil par la Kinect 56

Figure IV.8: Emissions de chaleurs qui proviennent essentiellement des yeux et de la bouche 57

Figure IV.9: Une scène réelle et l'image recueillie par la camera Kinect 57

Figure IV.10: Schéma résume la technologie Kinect.....58

Figure IV.11:*Kinect ASUS Xtion PRO LIVE RGB and Depth Sensor*.....59

Figure IV.12:*Kinect ASUS Xtion PRO LIVE RGB*……..……..60

Figure IV.13:*Schéma résumant l'ensemble des technologies Kinect*……61

Figure IV.14:*Image Original 1*……...…….63

Figure IV.15:*Segmentation 2D :a)segmentation des objets rouges, b) segmentation des objets vert, c)segmentation des objets bleu.*…….64

Figure IV.16:*Segmentation 3D :a)segmentation des objets rouge l'espace, b) segmentation des objets vert dans l'espace, c) segmentation des objet bleu dans l'espace*……...64

Figure IV.17:*la segmentation 3D de la 1 ère essaie*……...65

Figure IV.18:*Représentation 3D des objets colorés selon l'axe (X)*……..…..66

Figure IV.19:*Représentation 3D des objets colorés selon l'axe (Y)*……..…...67

Figure IV.20:*Représentation 3D des objets colorés selon l'axe (Z)*……...…….68

Figure IV.21:*Image Original 2*…….....…...69

Figure IV.22:*Représentation des objets segmenté en 3D*….....70

Figure IV.23:*Image Original 3*…….....…..73

Figure IV.24:*Représentation de la segmentation 3D*….....74

Figure IV.25:*Image quatrième essaie* …….....…..77

Figure IV.26:*Segmentation 2D :a)segmentation des objets rouges, b) segmentation des objets bleu*…….....……78

Figure IV.27:*la segmentation 3D de la quatrième essaie*….....78

Liste des tableaux

Tableau I.1 : *Tableau comparatif*.....12

Tableau IV.1: *Distances Des objets Prisent Manuellement (Exp1)*…….....63

Tableau IV.2: *Résultats des Distances Des objets Prisent Par Kinect (Exp1)*.....65

Tableau IV.3 : *Résultats d'erreur de la segmentation 3D suivants X(Exp1)*.....67

Tableau IV.4 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp1)*.....68

Tableau IV.5 : *Résultats d'erreur de la segmentation 3D suivants Z (Exp1)*……...…69

Tableau IV.6 : *Distances Des objets Prisent Manuellement (Exp2)*…….....…70

Tableau IV.7 : *Résultats des Distances Des objets Prisent Par Kinect (Exp2)*….....71

Tableau IV.8 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp2)*…….....…72

Tableau IV.9 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp2)*…….....73

Tableau IV.10 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp2)*…….....….73

Tableau IV.11 : *Distances Des objets Prisent Manuellement (Exp3)*…….....74

Tableau IV.12 : *Résultats des Distances Des objets Prisent Par Kinect (Exp3)*…….....…….75

Tableau IV.13 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp3)*…….....……76

Tableau IV.14 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp3)*….....77

Tableau IV.15 : *Résultats d'erreur de la segmentation 3D suivants Y (Exp3)*……...……77

Tableau IV.16 : *Distances Des objets Prisent Manuellement (Exp4)*…….....….78

Tableau IV.17 : *Résultats des Distances Des objets Prisent Par Kinect (Exp4)*…….....……79

Sommaire

Introduction générale 1
Chapitre I : Introduction au traitement d'image 2
I. Introduction 3
I.1 Notions générales sur l'image 3
I.1.1 Définition 3
I.1.2 Image numérique 3
I.1.2.1 Caractéristiques d'une image numérique 4
I.1.2.2 Définition du PIXEL 4
I.1.2.3 Distance entre 2 pixels 4
I.2 Pavages, maillage et connexité d'un pixel 4
I .2.1 Pavages du plan 4
I .2.2 Maillages 5
I .2.3 Connexité 5
I.3 La taille d'une image 6
I .4 La résolution 6
I.5 L'histogramme 7
I.6 Contours et textures 9
I.6.1 Luminance 9
I.6.2 Contraste 9
I.6.3 Balayage 9
I.6.4 Échantillonnage 9
I.6.5 Quantification 10
I.6.6 Le seuillage 10
I.6.7 Le bruit 11
I.7 Formats d'images 11
I.7.1 Définition 11

I. 7.2 Les différents formats d'image 11
I.7.3 Types d'images 12
I.8 Le Voxel 15
I.8.1 Différence entre "voxel" et "boxel".....16
I.9 Conclusion 17
Chapitre II : La Segmentation d'images RGB et ces applications **17**
II Introduction 19
II.1 Discussions 20
II.2 Stratégies de segmentation d'images couleur 20
II.2.1 Stratégie scalaire 21
II.2.2 Stratégie marginale 21
II.2.3 Stratégie vectorielle 20
II.3 Méthodes de segmentation d'images couleur basées régions 22
II.3.1 Méthodes de segmentation d'images couleurs par analyse spatiale 23
II.3.1 .1 La segmentation par division de régions 24
II.3.1 .2 La segmentation par croissance de régions 25
II.3.1 .3 La segmentation par fusion de régions 26
II.3.1.4 La segmentation par division-fusion de régions 26
II.3.2.1 Méthodes de segmentation d'images couleur par classification de pixels 28
II.3.2.2 Analyse d'histogrammes monodimensionnels 30
II.3.2.3 Isodata (Iterative Self-Organizing Data Analysis Techniques A) 31
II.3.3 Méthodes qui nécessitent la connaissance du nombre de classes de pixels.....31
II.3.3.1 Méthodes des Centre mobiles, nuées dynamiques, k-means 31
II.3.3.2 Méthode des Fuzzy c-means.....32
II.4 Conclusion.....32
Chapitre III : La reconstruction 3D 33
III Introduction.....34
III.1 La vision par ordinateur 34
III.2 Description d'un système de vision par ordinateur.....35

III.2.1 Définition 35
III.2.2 Classification des systèmes de vision par ordinateur 35
III.2.2.1 Les systèmes de vision bidimensionnelle 35
III.2.2.2 Les systèmes de vision tridimensionnelle 36
III.3 La reconstruction 3D 36
III.3.1 Les différentes approches de la reconstruction tridimensionnelle 36
III.4 Techniques de reconstruction 3D 37
III.4.1 La vision stéréoscopique 38
III.4.1.1 Principe 38
III.4.1.2 L'appariement 39
III.4.1.3 La triangulation 39
III.4.2 Reconstruction à partir de géométrie epipolaire 40
III.4.3 Technique base sur les ombrages (Shapefrome shading) 42
III.4.4 Représentation par détection de déformation de formes 43
III.4.5 Méthodes basées sur l'extraction de squelette 44
III.4.6 La reconstruction par des jonctions caractéristiques 44
III.4.7 La reconstruction à partir du mouvement (Shape From Motion) 45
III.4.7.1 L'approche basée sur la mise en correspondance 46
III.4.7.2 L'approche du flot optique 46
III.4.8 Méthode par spacecarving (creusage d'un volume 3D) 47
III.5 Conclusion 48
Chapitre IV : Application et Résultats 49
IV. Introduction 50
IV.1 La Kinect 50
IV.2 Principe de Fonctionnement 51
IV.2.1 Concept 51
IV.2.2 La camera RGB (Red Green Blue) 52
IV.2.2.1 Le capteur photographique 52
IV.2.2.2 Fonctionnement du capteur 54

IV.2.3 Le 3D Depth Sensor 56
IV.2.3.1 Comment fonctionne un capteur CMOS infrarouge 56
IV.3 Description de la méthode utilisée pour notre application.....59
IV.3.1 Le Matériel utilisé 59
IV.3.1.1 A propos de la Kinect ASUS Xtion PRO LIVE RGB and Depth Sensor.....59
IV.3.1.2 Principe 59
IV.4 L'algorithme.....62
IV.5 Résultats et discutions.....63
IV.5.1 Les Expériences 63
IV.6 Conclusion 80
Conclusion générale.....81
Bibliographie

Introduction générale

Introduction générale

Dans ce mémoire nous nous intéressons à la segmentation des images RGBD (Red, Green, Blue and Depth), on utilisant un programme qui permet la segmentation par histogramme. Pour ce faire, nous avons divisé ce travail en quatre chapitres. Dans le premier chapitre, nous présenterons des généralités sur le traitement d'image parlons de quelques notions de base qui serons utilisé par la suite dans notre travail. Nous exposerons dans le deuxième chapitre La segmentation d'image RGB et ces applications présenter les différents méthodes de segmentations les plus utilisé dans le domaine d'imagerie. Dans le troisième chapitre, nous aborderons La reconstruction 3D d'un point de vue théorique et pratique exploité par les programmes. Par la suite le chapitre quatre seras traduit le coté pratique de notre travail de recherche en basont sur plusieurs expériences pour testent l'efficacité du programme proposé.

L'évolution d'image analogique jusqu'à l'image numérique, a changé nos aidées sur le monde d'imagerie. L'informatisation d'image a aussi changé beaucoup de principes et de concepts. Donc avoir des connaissances sur le traitement d'image, permis de nous éclairer.

Le but de la segmentation d'images a pour but l'extraction de l'information caractéristique contenue dans une image. Qui donne une description structurelle de l'image ou de toute structure de donnée (Niveaux de gris, couleur, forme, texture, ….).

Quelque définition concernant la vision par ordinateur et les systèmes de vision et la classification de ces systèmes serons données puis on parle de la reconstruction 3D en citant quelque méthode utilisée pour reconstruire des objets 3D.

On présente les résultats d'implémentation de la méthode utilisé la segmentation RGBD par la méthode de détection des objets par le billé d'histogramme et discuté ces résultats pour arrivé a des conclusions rigoureuses.

Chapitre I

Introduction au traitement d'image

Chapitre I : Introduction au traitement d’image

- **INTRODUCTION.**
- **Image numérique.**
- **Contours et textures**
- **Le seuillage**
- **Formats d'images**
- **Le Voxel**
- **CONCLUSION.**

I. Introduction:

Depuis l'image analogique jusqu'à l'image numérique, l'image numérique a changé nos aidées sur le monde d'imagerie. L'informatisation d'image a changé beaucoup de principes et de concepts. Donc avoir des connaissances sur le traitement d'image, permis de nous éclairer et de nous guidé dans ce mémoire, d'où l'intérêt de ce premier chapitre qui contient des brèves définitions, Caractérisation et concepts qui peuvent être utiles

I.1 Notions générales sur l'image :

I.1.1 Définition :

L'image est une représentation d'une personne ou d'un objet par la peinture, la sculpture, le dessin, la photographie, le film…. C'est aussi un ensemble structuré d'informations qui, après affichage sur l'écran, ont une signification pour l'œil humain. Elle peut être décrite sous la forme d'une fonction I(x, y) de brillance analogique continue, définie dans un domaine borné

- Les x et y sont les coordonnées spatiales d'un point de l'image.
- I est une fonction d'intensité lumineuse et de couleur.

Sous cet aspect, l'image est inexploitable par la machine, ce qui nécessite sa numération **[03].**

I .1.2 Image numérique

Une image numérisée est un ensemble de pixels (Picture élément), à chacun desquels est attribué un niveau de gris, de 0 (noir) à 255 (blanc). Le nombre $256 = 2^8$ traduit le codage sur 8 bits en binaire, par exemple : 10010101.

Une image numérisée en niveaux de gris n'est donc rien d'autre qu'une fonction définie sur une grille discrète et prenant des valeurs discrètes, c'est-à-dire un tableau (matrice) de nombres (binaires). Une telle image contient donc une quantité d'information énorme. Toutefois, de par sa définition même, elle va présenter des difficultés géométriques : courbes et droites sont mal reproduites, il y a des effets de pixellisation **[1]**. La Figure I.1 représente une image numérique.

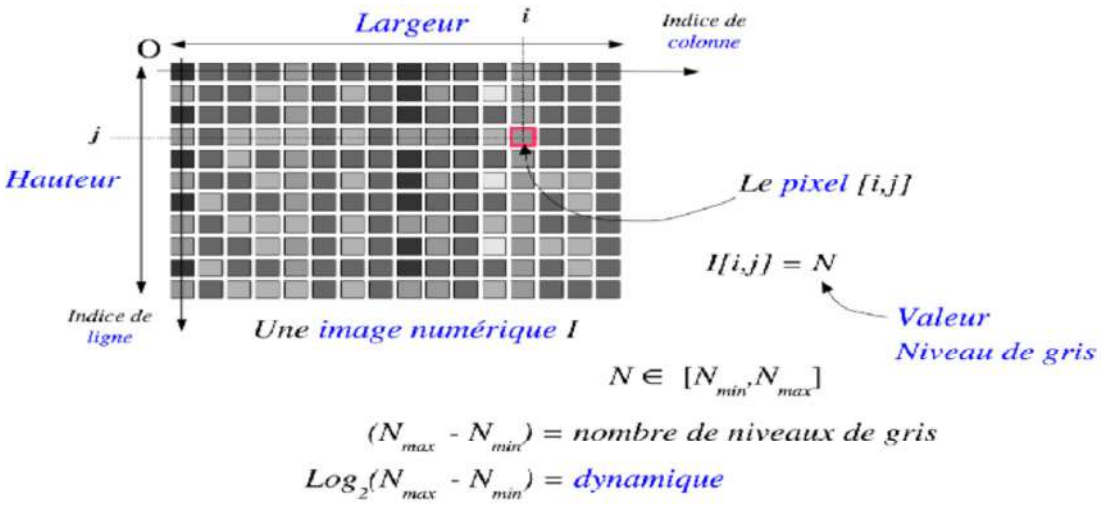


Figure I.1 représentation d'une image numérique

I.1.2.1 Caractéristiques d'une image numérique :

I.1.2.2Définition du PIXEL :[6]

PIXEL est l'abréviation de Picture Elément. C'est le plus petit point de l'écran que l'on peut allumer ou éteindre. Il possède deux caractéristiques :

• Position (coordonnée X,Y)

• niveaux de gris ou Couleur 1, 4, 8, 16 ou 24 bits

I.1.2.3Distance entre 2 pixels [04] :

L'image est caractérisé par un couple de coordonnée (X, Y).

Pour deux pixels P (Xp, Yp) et Q (Xq, Yq) :

-Distance de manathan (city block): D (P, Q)= | Xp-Xq | + | Yp-Yq | ….. (I.1).

-Distance de l'échiquier : D (P, Q)= Max (| Xp-Xq |, | Yp-Yq |) …… (I.2).

I.2 Pavages, maillage et connexité d'un pixel :

I .2.1 Pavages du plan : [6]

Un pavage du plan est une partition du plan en cellules élémentaires (pixels).

Il n'existe que 3 pavages réguliers du plan :

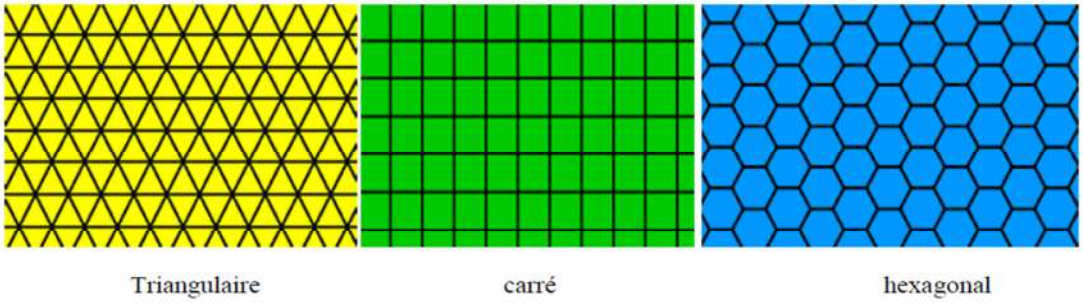


Figure I.2 Les types de pavages

I .2.2 Maillages :[6]

A tout pavage du plan on peut associer un graphe où les sommets (nœuds) représentent les cellules élémentaires, et où les arêtes représentent la relation d'adjacence entre les cellules (2 cellules sont adjacentes si elles ont un côté en commun). Un tel graphe est un Maillage du plan.

-Maillage carré : qui correspond à la réalité physique des capteurs d'image.

-Maillage hexagonal: est une référence pour le domaine de la morphologie mathématique.

-Maillage triangulaire: est plus confidentiel.

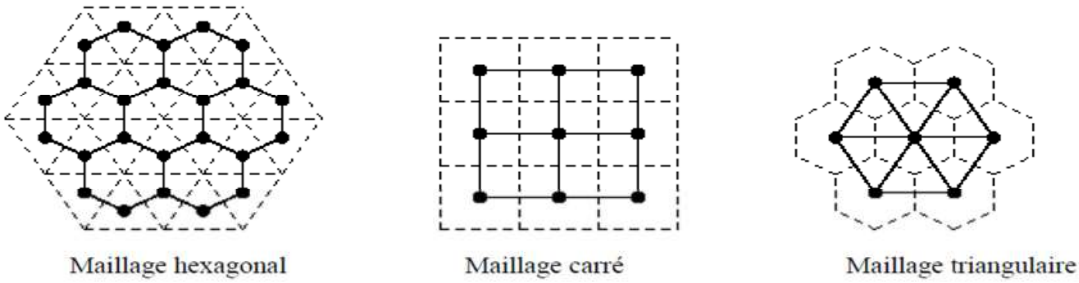


Figure I.3 Les types de Maillage

I .2.3 Connexité[6]

Les relations topologiques dans les images discrètes sont définies à partir de la relation de connexité induite par le graphe du maillage (X,S), où X représente les sommets et S les arêtes.

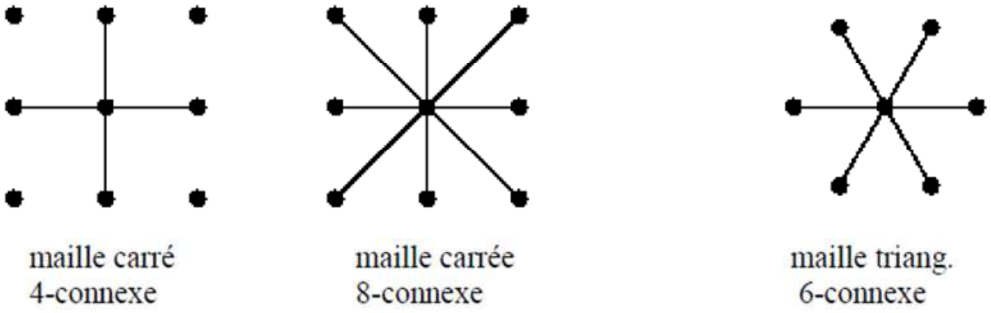


Figure I.4 Les types de connexité

I.3 La taille d'une image :

Pour connaître la taille d'une image, il est nécessaire de compter le nombre de pixels que contient l'image, cela revient à calculer le nombre des cases du tableau, soit la hauteur de celui-ci que multiplie sa largeur. La taille de l'image est alors le nombre des pixels que multiplie la taille (en octet) de chacun de ces éléments. **[04]**

Exemple :

Pour une image de 240 X 420

Nombre de pixels : 240 X 420 = 100800

Taille de chaque pixel : 24 bits / 8 = 3 octets

Le poids de l'image est ainsi égal à :

100800 X 3 = 302.400 et Égal : 302.400/1024 =295 Ko

I .4 La résolution :[6]

La résolution d'une image est le nombre de pixels contenus dans l'image par unité de longueur. Elle s'exprime le plus souvent en ppp (point par pouces) ou en dpi (dots per inch) parfois en point par cm. La résolution définit la netteté et la qualité d'une image. Plus la résolution est grande (c'est-à-dire plus il y a de pixels dans une longueur de 1 pouce), plus votre image est précise dans les détails. Il ya deux types de résolution :

Résolution spatiale : Le plus petit détail discernable



Figure I.5 : Résolution spatiale

Résolution tonale :(de tons de gris) : Le plus petit changement discernable

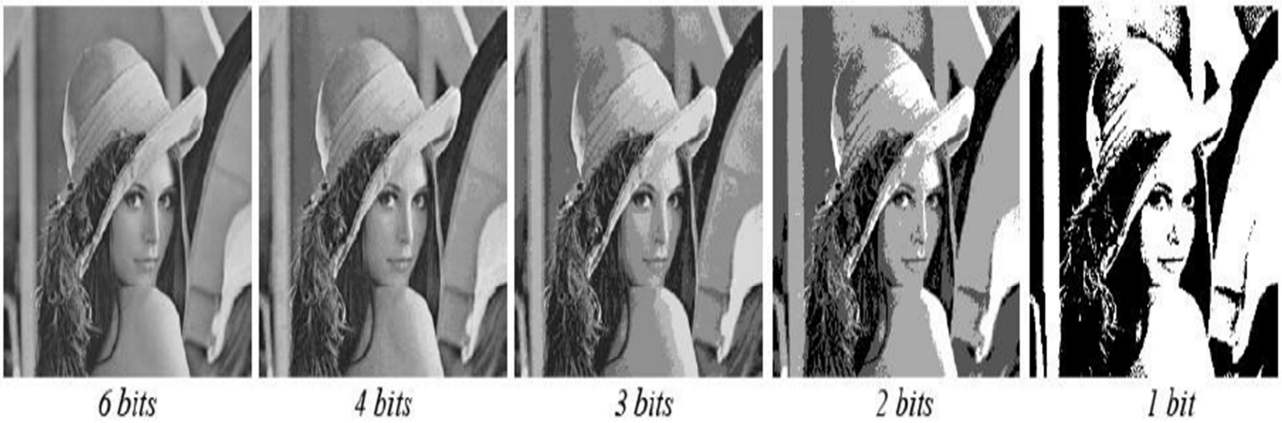


Figure I.6 : Résolution tonale

I.5 L'histogramme

L'histogramme est la représentation graphique de la proportion de pixels d'une image par gamme de luminosité. Au-delà de cette définition un peu barbare, l'histogramme permet de répondre concrètement aux questions suivantes : « Combien de pixels de mon image sont-ils dans les tons foncés (ou la haute lumière) ? », ou encore « Quelle est la luminosité majoritaire dans mon image? ». Du fait qu'il s'agit d'une représentation graphique, l'histogramme permet de répondre à ces questions sans calcul complexe.

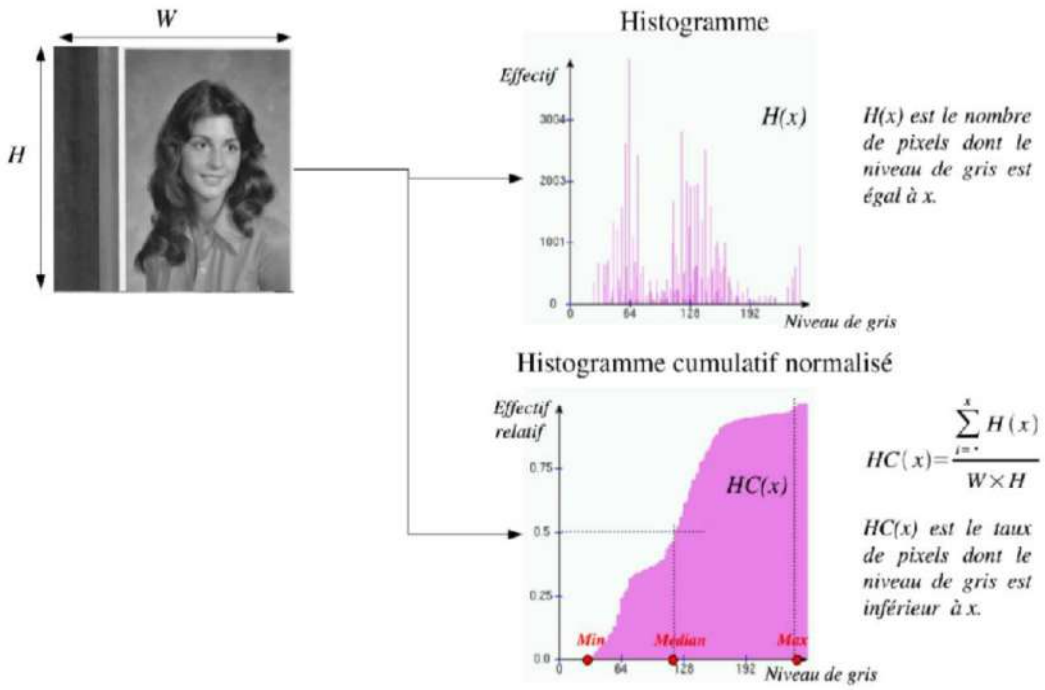


Figure I.7 : représentation d'histogramme [06]

Il existe deux types d'histogramme :

a. Normalisation :

La normalisation d'histogramme, ou expansion de dynamique, est une transformation affine du niveau de gris des pixels de telle sorte que l'image utilise toute la dynamique de représentation.

$$f\,new[x,y] = \frac{(2^d - 1)}{(N_{max} - N_{min})}.(f[x,y] - N_{min}) \dots\dots\dots\dots\dots\dots\dots\dots (I.3)$$

Avec, d : dynamique ; N_{max} : la plus grande valeur dans l'image et N_{min} : la plus petite valeur dans l'image.

b. Égalisation :

L'égalisation d'histogramme est une transformation des niveaux de gris dont le principe est d'équilibrer le mieux possible la distribution des pixels dans la dynamique (Idéalement, on cherche à obtenir un histogramme plat).

$$f\,new[x,y] = (2^d - 1).\frac{HC(f[x,y])}{wh} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (I.4)$$

Avec, d : dynamique **;** (w, h) : dimension de l'image et HC : histogramme cumulé.

I.6 Contours et textures :

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative. Les textures décrivent la structure de ceux-ci. L'extraction de contour consiste à identifier dans l'image les points qui séparent deux textures différentes. [07]

I.6.1 Luminance :

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet.

Une bonne luminance se caractérise par [08] :

a. Des images lumineuses (brillantes);

b. Un bon contraste : il faut éviter les images où la gamme de contraste tend vers le blanc Ou le noir; ces images entraînent des pertes de détails dans les zones sombres ou lumineuses.

c. L'absence de parasites.

I.6.2 Contraste :

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images. [08]

Si L1 et L2 sont les degrés de luminosité respectivement de deux zones voisines A1 et A2 d'une image, le contraste c'est défini par le rapport :

$$C = (L1-L2)/(L1+L2) \quad \text{(I.5)}$$

I.6.3 Balayage :

Analyse d'une image ligne à ligne par l'intermédiaire d'un rayon lumineux, le plus souvent dans le cadre d'une numérisation (scanner). **[08]**

I.6.4 Échantillonnage :

L'échantillonnage est une étape fondamentale qui doit tenir compte du contenu informationnel pertinent de l'image à analyser, elle est le procédé de discrétisation spatiale d'une image consistant à associer à chaque zone rectangulaire R(x,y) d'une image continue une unique valeur I(x,y). **[08]**

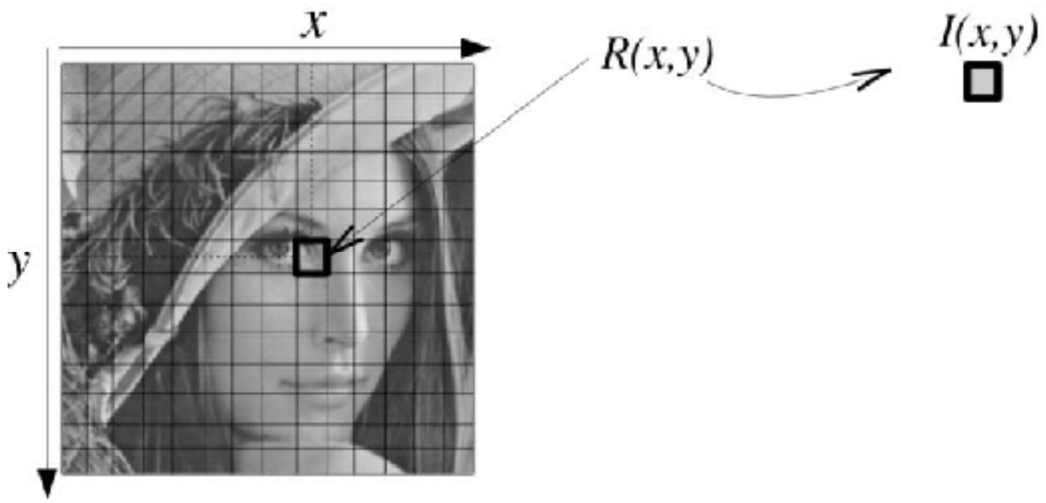


Figure I.8 : échantillonnage d'image

I.6.5 Quantification :

Transformation d'une amplitude à valeurs continues en une amplitude qui ne peut prendre qu'un nombre fini de valeurs distinctes. Pour une image à 256 niveaux de gris, chaque amplitude est codée sur un octet (8 bits). **[08]**

I.6.6 Le seuillage :

Une image numérique, pour pouvoir être exploitée a généralement besoin d'être simplifiée, c'est le but du seuillage. Ce dernier consiste à transformer l'image codée sur 6, 8 ou 16 bits, en une image binaire ou les pixels à 1 correspondent aux objets et les pixels à 0 au fond de l'image. A la différence des différents traitements de l'image numérique, le seuillage est un passage obligatoire pour toute analyse morphologique ultérieure. Le seuillage permet de sélectionner les parties de l'image qui intéressent l'opérateur, par exemple 2 types de grains (blancs et sombres) dans un mélange. On peut donc, par exemple, attribuer à tous les pixels de l'image numérique qui ont un niveau de gris compris entre deux valeurs i1 et i2, choisies par l'opérateur, la valeur 1; à tous les autres pixels est attribuée la valeur 0.

Après seuillage, les parties de l'image sélectionnées seront traduites en noir et blanc. L'image, digitalisée par l'ordinateur (0 et 1), est appelée image binaire. Cette dernière, tout comme l'image numérique contient des informations superfétatoires, qu'il convient d'éviter, ou masquées qu'il faut révéler. Les traitements suivants permettent de modifier l'image binaire à ces fins.

Il n'est pas en général nécessaire de s'intéresser à tous les paramètres de base relatifs à une image, car seules certaines propriétés sont utiles. L'idée est donc de transformer l'image binaire pour n'accéder qu'aux informations qui nous préoccupent. De plus, le seuillage de l'image peut faire apparaître des défauts dus à l'irrégularité de la surface ou à d'autres causes externes. La morphologie mathématique permet de transformer ces images en « tout ou rien » grâce à des éléments structurant.

I.6.7 Le bruit :

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.

I.7. Formats d'images :

I.7.1 Définition :

Un format d'image est une représentation informatique de l'image, associée à des informations sur la façon dont l'image est codée et fournissant éventuellement des indications sur la manière de la décoder et de la manipuler. **[01]**

I. 7.2 Les différents formats d'image :

Un format d'image est une représentation informatique de l'image, associée à des informations sur la façon dont l'image est codée et fournissant éventuellement des indications sur la manière de la décoder et de la manipuler.

La plupart des formats sont composés d'un en-tête contenant des attributs (dimensions de l'image, type de codage, etc.), suivi des données (l'image proprement dite). La structuration des attributs et des données diffère pour chaque format d'image. De plus, les formats actuels intègrent souvent une zone de métadonnées (Meta data en anglais) servant à préciser les informations concernant l'image comme : la date, l'heure et le lieu de la prise dévoue, les caractéristiques physiques de la photographie (sensibilité ISO, vitesse d'obturation, usage du flash.

Tableau I.1 : Tableau comparatif

	Type (matriciel/ Vectoriel)	Compressions des données	Nombre de couleurs supportées	Affichage progressif	Animation	Transparence
JPEG (Joint Photographic Experts Group)	matriciel	Oui, réglable (avec perte)	16 millions	Oui	Non	Non
JPEG2000	matriciel	Oui, avec ou sans perte	4 milliards	Oui	Oui	Oui
GIF (Graphics Interchange Format)	matriciel	Oui, Sans perte	256 maxi (palette)	Oui	Oui	Oui
PNG (Portable Network Graphics)	matriciel	Oui, Sans perte	Palettisé (256 couleurs ou moins) ou 16 millions	Oui	Non	Oui (couche Alpha)
TIFF (Tagged Image File Format)	matriciel	Compression ou pas avec ou sans pertes	de monochrome à 16 millions	Non	Non	Oui (couche Alpha)
SVG (Scalable Vector Graphics)	vectoriel	compression possible	16 millions	ne s'applique pas	Oui	Oui (par nature)

I.7.3 Types d'images [2]

a- Les images (matricielles) ou bitmap.

Une image matricielle (ou bitmap) est une image constituée d'un ensemble de points : les pixels. Chaque point porte des informations de position et de couleur. Format d'images bitmap : BMP, PCX, GIF, JPEG, TIFF.

Les photos numériques et les images scannées sont de ce type.

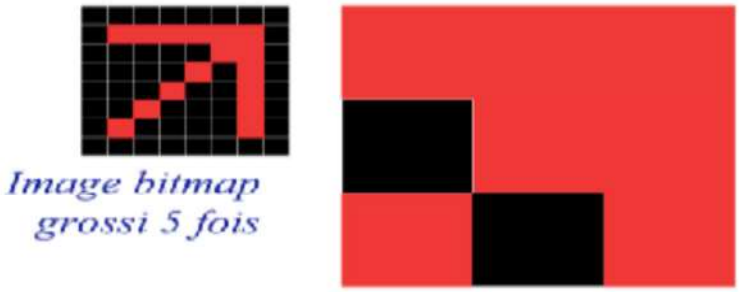


Figure I.9 : exemple d'image matricielle

b- Les images vectorielles :

Les images vectorielles sont composées de formes géométriques qui vont pouvoir être décrites d'un point de vue mathématique. Par exemple une droite sera définie par 2 points, un cercle par un centre et un rayon. Le processeur est chargé de "traduire" ces formes en informations interprétables par la carte graphique (images Word, Publisher, CorelDraw - format WMF, CGM, etc.)

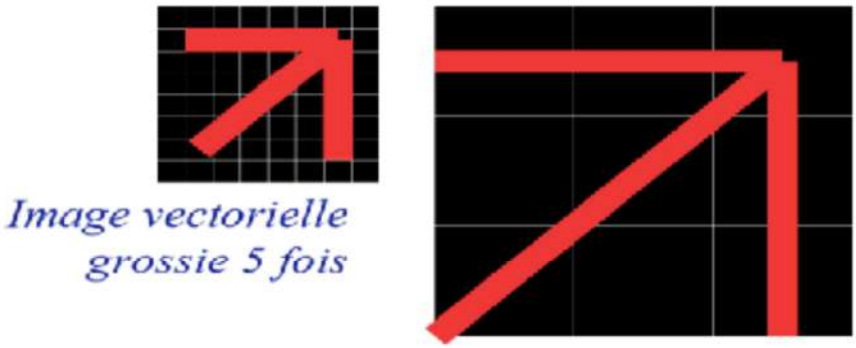


Figure I.10 : exemple d'image vectoriel

c- Images en couleurs :

Une image couleur correspond à la synthèse additive de 3 images, rouge, vert et bleu. Chaque pixel est donc codé sur 3×Nbits.

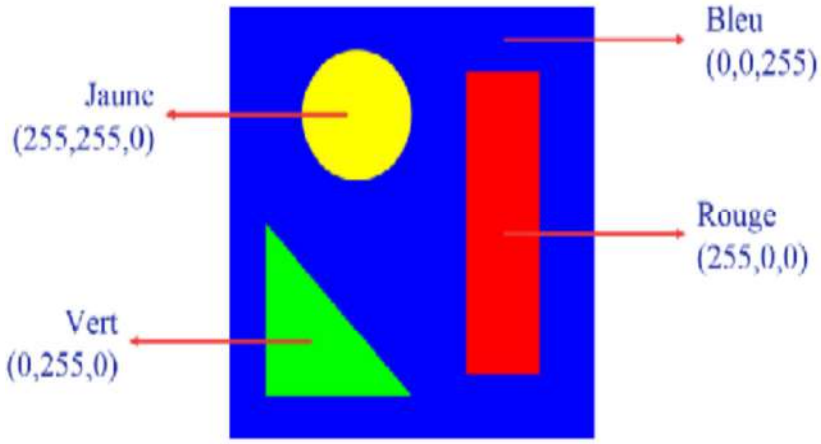


Figure I.11 : exemple d'image en couleur

d- Images binaire :

Dans une image binaire, les pixels sont représentés par deux états logiques 0 (noir) et 1 (blanc). C'est un codage de l'image sur 1 bit.

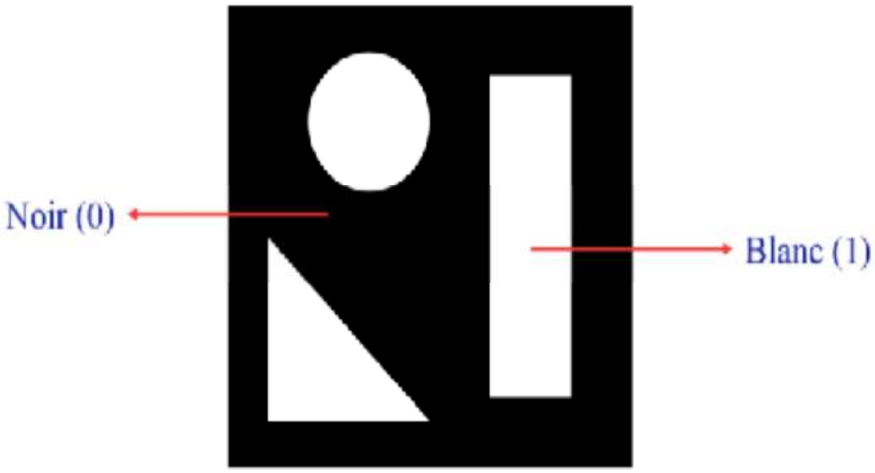


Figure I.12 : exemple d'image binaire

e- Images en niveau de gris :

Chaque pixel est codé sur N bits, ce qui lui confère des valeurs entières comprises entre 0 (noir) et (2^n-1) (blanc).

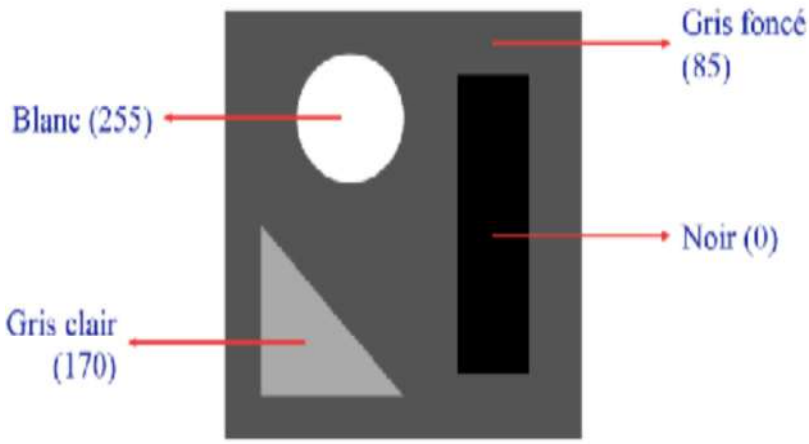


Figure I.13 : exemple d'image en niveau de gris

I.8 Le Voxel :

Le voxel (contraction de « volumétrique pixel ») est un pixel en 3D. Le voxel est un terme nettement plus souvent utilisé que son synonyme, le boxel. Il consiste à stocker une information colorimétrique avec ses coordonnées spatiales, voire temporelles, de positionnement ainsi que, facultativement, une taille relative à l'unité utilisée ou d'autres informations telles qu'une matière… Ils s'inscrivent plus généralement dans des espaces matriciels bien que les espaces vectoriels leur soient favorables. On notera que leurs coordonnées spatiales peuvent être polaires.

Leur point faible (par rapport à la 3D vectorielle) est de nécessiter énormément de ressources, tant pour leur stockage que pour leur rendu. Cependant, des algorithmes de compression traditionnels à l'imagerie numérique ou l'informatique peuvent être employés sans difficulté.

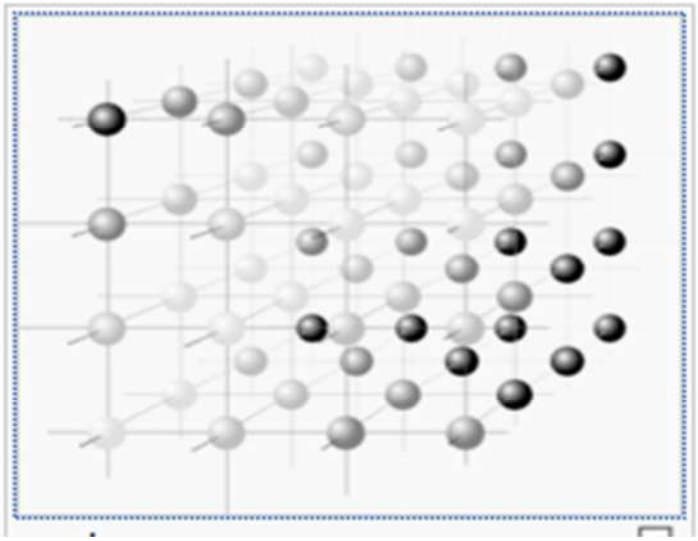


Figure I.14 Schématisation d'une matrice de voxels en nuances de gris.

I.8.1 Différence entre "voxel" et "boxel" :

Le terme "boxel" dérive du terme anglais "box", qui signifie "boite". Ils constituent une sous-catégorie des voxels. Le terme "voxel" est donc plus généraliste que le terme "boxel" et l'on distinguera alors l'aspect vectoriel des voxels en opposition à l'aspect matriciel des boxels.

Un boxel n'est rien d'autre qu'un pixel volumique matriciel. Sa représentation basique se fera donc sous la forme d'un cube au sein d'une matrice (i.e., d'une grille) tridimensionnelle (i.e., en 3D). Cependant, diverses techniques de rendu permettent d'en arrondir les angles selon la nature des boxels adjacents (à la manière des métaballs). Ils sont généralement associés à des octrees pour leur manipulation et leur stockage (bien que ceci ne constitue en rien une obligation !).

Néanmoins, la nuance entre ces deux termes n'est pas réellement factuelle. En effet, la plupart des moteurs utilisant des voxels utilisent en réalité des boxels, et comme les boxels sont des voxels, on utilisera plutôt le terme de "boxel" dans le langage usuel pour faire référence à une méthode de traitement matricielle en opposition aux techniques vectorielles.

Cette différence n'est de fait pas réellement pertinente car les diverses techniques (de stockage, de manipulation, de rastérisation, de rendu, ...) sont souvent mixées et cumulées.

Par exemple, rien n'empêche de rendre graphiquement un voxel vectoriel avec les techniques utilisés par les boxels, c'est-à-dire, dans un espace matriciel via l'utilisation du Wu-

Pixeling appliqué aux espaces matriciels tridimensionnels. Ceci consiste juste à transposer les coordonnées vectorielles du voxel dans un espace matriciel constitué de boxels via une synthèse additive proportionnelle de la couleur des boxels adjacents.

Le voxel est utilisé pour la représentation d'espaces 3D, en physique ou en imagerie médicale 3D par traitement numérique de coupes 2D issues des machines d'investigation (Scanner, IRM ...).

Voir par exemple le scanner à rayons X (ou CT).

I.9 Conclusion :

Les images numériques sont de plusieurs types : couleurs, noir et blanc, niveaux de gris et couleurs. Celles en couleurs peuvent être représentées sur des différentes espaces.

Nous avons décrit dans ce chapitre quelques notions sur l’analyse d’image, numérisation (balayage, échantillonnage et quantification).

Chapitre II

La segmentation d’image RGB et ces application

Chapitre II : La segmentation d'image RGB et ces application

- **INTRODUCTION.**
- **Discussions.**
- **Stratégies de segmentation d'images couleur**
- **Méthodes de segmentation d'images couleur basées régions**
- **CONCLUSION**

II Introduction :

L'analyse d'images a pour but l'extraction de l'information caractéristique contenue dans une image. Cette analyse donne une description structurelle de l'image, qui peut prendre la forme d'une image ou de toute structure de donnée (Niveaux de gris, couleur, forme, texture, ….). L'analyse d'images fait appel à la segmentation ou l'on va associer à chaque Pixel un label.

La segmentation consiste à créer une partition de l'image en région homogènes, afin d'en extraire des informations particulière. La segmentation dépend de la nature de l'image et de l'objectif recherché. Par conséquent, il convient d'analysé les images avant de définir une méthode de segmentation appropriées.

A la différence du prétraitement qui transforme l'image couleur en une autre image couleur, la segmentation transforme l'image couleur soit en une image binaire (0 : noire, 1 : blanc) contour/non contour, soit en une image de régions pour laquelle tous les pixels d'une même région sont affectés d'une même étiquette.

Dans la littérature, on trouve une multitude de méthodes de segmentation, mais aucune ne permet d'obtenir des résultats satisfaisants pour toutes les images. Le choix de la méthode dépend souvent des connaissances à priori sur l'image et repose sur les différents aspects suivants [09] :

- La présence ou non de texture, bruit.
- Le choix de l'espace couleur.
- Le but recherché (isoler des régions pertinentes, extraire des objets,…)
- Le choix entre extraction des contours ou extraction des régions.

De manière générale, lorsqu'il s'agit d'images couleur comme dans notre cas, les méthodes de segmentation basées contours sont proscrites puisqu'une texture ne comporte en général pas de frontières franches. Les approches régions sont plus adaptées [11, 10]. Nous nous intéresserons beaucoup plus aux méthodes d'extraction de régions.

Dans ce chapitre, un état de l'art de toutes les méthodes de segmentation basées régions sera fait.

II.1 Discussions :

Toutes les méthodes de caractérisations d' image couleur qu'on a vu sont les plus Couramment utilisées. Il existe cependant d'autres méthodes et attributs moins. Nous dirons que les caractéristiques d'une« bonne » méthode seront caractérisées par :

- La rapidité (temps de calcul).
- Moins d'espace mémoire, c'est à dire moins gourmande en espace mémoire.
- Spatiale et couleur, c'est à dire elle tient compte des interactions spatiales et couleur.
- Inter et intra-composantes, c'est à dire l'analyse de la texture couleur se ferait à la foisau niveau des différentes images composantes

Le choix de la méthode à utiliser est un point très délicat, il dépend de plusieurs facteurs comme la régularité de la texture, les couleurs de la texture, le temps de réponse et la mémoire de stockage des attributs

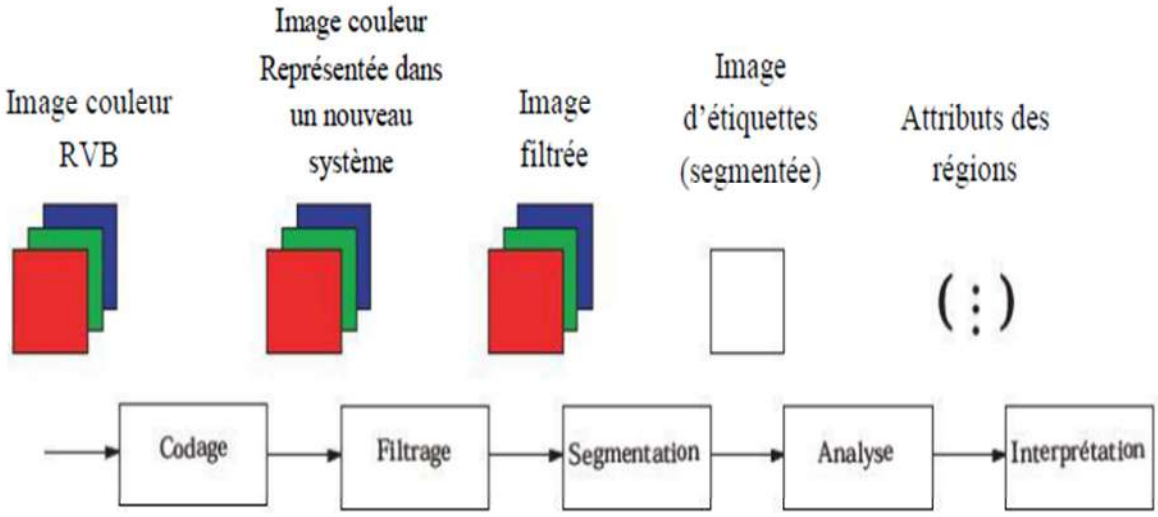


Figure II.1 Chaine de traitement d'une image couleur

II.2 Stratégies de segmentation d'images couleur

Cette section présente d'abord les différentes stratégies qu'il est possible d'adopter pour traiter les images couleur (multi-composantes). Puis sont abordées les méthodes de segmentation basées régions. Il existe trois stratégies possibles pour la segmentation des images couleur [09, 12] :

II.2.1 Stratégie scalaire :

Il s'agit de passer d'une image couleur à une image monochrome qui peut être fait par la quantification ou par la transformée de Karhunen-loéve (ACP). Toutes les méthodes de segmentation monochrome sont alors possible ; il est bien sûr recommandé comme dans tous les cas, de choisir la plus adéquate à l'application choisie (Figure II.2)

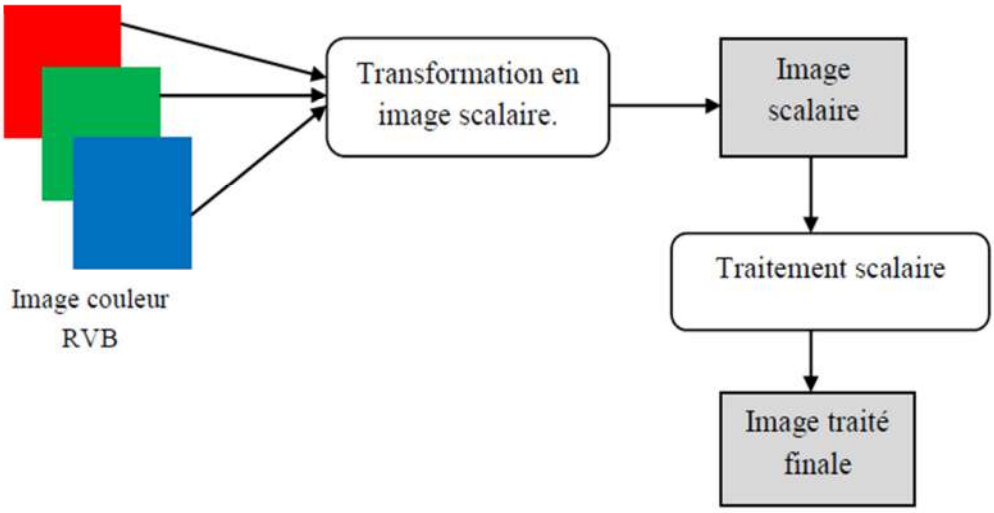


Figure II.2 Stratégie scalaire.

II.2.2 Stratégie marginale :

Elle consiste à traiter les images composantes séparément, les trois résultats sont ensuite fusionnés (Figure II.3).

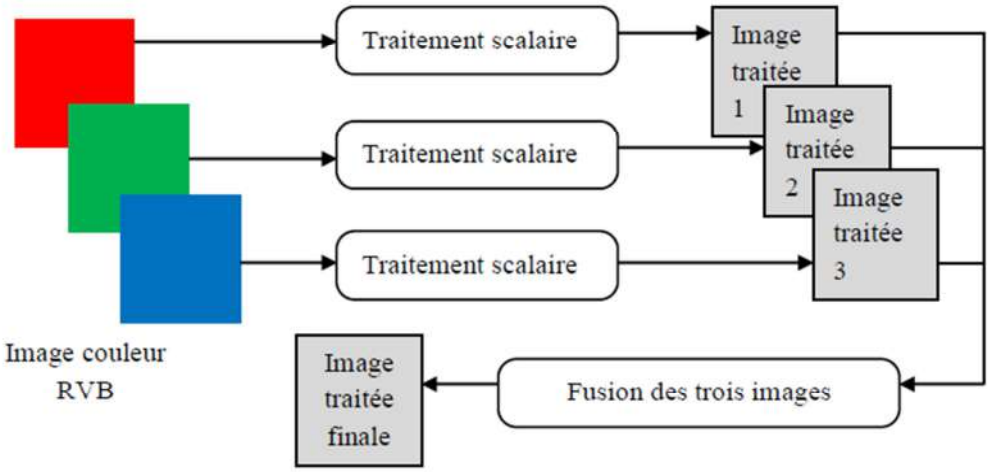


Figure II.3 Stratégie marginale.

II.2.3 Stratégie vectorielle :

Elle consiste à traiter chaque pixel comme un vecteur dans l'espace couleur choisi. Les primitives sont extraites en tenant compte des trois composantes à la fois (Figure II.4)

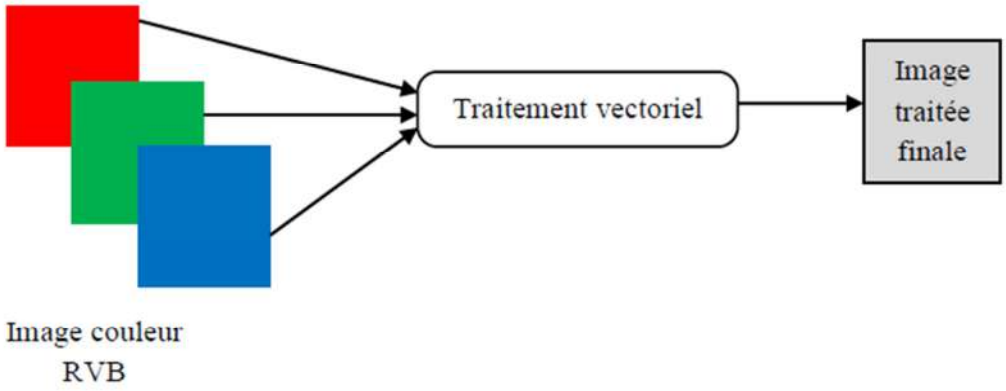


Figure II.4 Stratégie vectorielle.

Il est important de noter qu'il existe une autre stratégie dite semi-vectorielle, cette stratégie est une hybridation de la stratégie marginale et vectorielle [12].

II.3 Méthodes de segmentation d'images couleur basées régions :

La segmentation en régions d'une image *I* a pour but de regrouper des pixels connexes ayant des attributs similaires (couleur, texture…), afin de constituer des régions *Ri* homogènes et cohérentes, ces régions étant dans la mesure du possible liées à l'information utile que nous voulons extraire de l'image initiale *I*.

Dans les images couleur, les attributs utilisés sont issus des méthodes d'analyse de texture couleur. Les pixels de chaque région doivent vérifier les attributs utilisés pour respecter le critère d'homogénéité, et ils doivent aussi respecter le critère de connexité.

L'homogénéité d'une région *Ri* est définie par un prédicat d'uniformité, noté *pred(Ri)*. Ce prédicat est vrai si *Ri* est homogène, faux dans le cas contraire. Soit *R* la région qui représente l'image entière *I*, les sous-régions *Ri* doivent respecter les cinq conditions suivantes [13, 14] :

- R= $\cup i$ R*i*: cette condition implique que chaque pixel de l'image doit appartenir à une région *Ri* et que l'union de toutes les régions correspond à l'image entière.
- Riest constituée de pixels connexes pour tout *i* : cette condition est relative à la structure des régions. Elle définit une région comme un sous-ensemble de pixels connexes.
- Pred (R*i*) = vr ai pour tout *i :*cette condition exprime que chaque région doit respecter vun prédicat d'uniformité.
- *Ri* $\cap$*Rj*= Øpour tout *i* et *j*, $i \neq j$: cette condition exige que les régions doivent être disjointes, c'est à dire bien séparées
- Pred (R*i*$\cup$*Rj*) = faut pour tout $i \neq j$, *Ri* et *Rj*étant adjacentes dans *R :* cette dernière condition implique la non réalisation de ce même prédicat pour toute réunion de deux régions adjacentes.

Le résultat de la segmentation est une image mono-composante dans laquelle on attribue à chaque pixel une étiquette correspondant à la région à laquelle il appartient.

Les méthodes de segmentation basées régions peuvent être divisées en deux grandes familles [04] suivant que les pixels sont analysés dans le plan image (spatial) ou dans l'espace couleur (classification).

- Méthodes de segmentation d'images couleurs par analyse spatiale.
- Méthodes de segmentation d'images couleurs par classification de pixels.

Ces deux familles de méthodes de segmentation d'images couleurs représentent une liste de méthodes non exhaustive, car on trouve dans la littérature une grande diversité de méthodes et de manières à les classifier.

Enfin, il est important de noter que quelques méthodes font coopérer les deux familles précédentes.

II.3.1Méthodes de segmentation d'images couleurs par analyse spatiale :

Ce type de méthodes se base sur l'analyse des pixels directement dans le plan image en parcourant l'image afin de regrouper les pixels connexes respectant le prédicat utilisé pour construire les régions. Elles privilégient essentiellement l'interaction spatiale entre les pixels.

Sachant que la décision d'associer les pixels afin de construire les régions est commune pour toutes les méthodes. Cette décision est binaire, le prédicat utilisé est donc soit vrai (1 logique) soit faux (0 logique) comme on l'a vu précédemment [09].

Nous pouvons distinguer quatre approches [14], et qui sont :

- La segmentation par division de régions,
- La segmentation par croissance de régions,
- La segmentation par fusion de régions,
- La segmentation par division-fusion de régions.

Les différences entre les méthodes de ce type proviennent principalement des critères utilisés par les prédicats qui sont par exemple [09] :

- La distance entre les couleurs de deux pixels (généralement la distance euclidienne).
- Un critère fondé sur l'analyse de la texture couleur qu'on utilisera dans notre cas.
- Un critère d'homogénéité couleur de la région en cours de reconstruction.

II.3.1 .1 La segmentation par division de régions :

L'approche segmentation par division de régions consiste à diviser l'image originale en régions homogènes au sens d'un critère donné.

Ce processus est récursif et considère que la région initiale (*R*) correspond à l'image à segmenter (*I*). Si une région ne respecte pas un prédicat d'homogénéité, elle est divisée en quatre sous-régions de taille égale (structure quadtree). Chaque sous-région est ensuite analysée. L'algorithme récursif s'arrête lorsque toutes les régions respectent le prédicat d'homogénéité. A cause de l'utilisation de la structure quadtree, cette méthode est plutôt adaptée à des images carrées ayant un nombre de lignes et de colonne égale à une puissance de deux, et dans lesquelles les régions sont de forme rectangulaire. D'autre part, cette méthode a tendance à faire apparaitre des effets de blocs [14] (Figure II.5) :

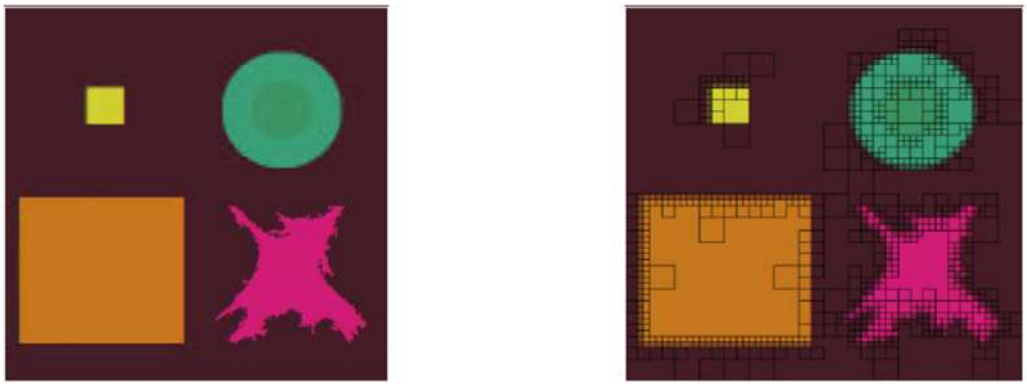


Figure II.5 Segmentation de l'image (a) par la méthode division de régions en donnant l'image segmentée (b).

II.3.1 .2 La segmentation par croissance de régions :

Cette approche consiste à explorer l'image à partir de regroupements initiaux de pixels connexes appelés germes, et à faire croitre ceux-ci sous une double contrainte d'adjacence et d'homogénéité. Les germes sont sélectionnés de manière déterministe ou aléatoire par prétraitement de l'image. De manière itérative, de nouveaux pixels sont ajoutés aux germes jusqu'à ce que la région en cours de construction ne respecte plus le ou les prédicats mis en place. La plupart de ces algorithmes parcourent l'image selon un ordre qui peut influencer le résultat de la segmentation. Le parcours le plus fréquent consiste à partir d'un germe en haut à gauche de l'image pour se diriger vers le bas à droite de l'image.

L'avantage de cette approche est de préserver la forme de chaque région de l'image. Cependant, une mauvaise sélection des germes ou un choix du prédicat mal adapté peuvent entrainer des phénomènes de sous-segmentation ou de sur-segmentation [14].

Plusieurs auteurs tentent d'améliorer les résultats de ce type d'approche en se focalisant sur les germes choisis pour limiter les phénomènes de sous-segmentation ou de sur-segmentation. Pour cela, ils ont opté pour les méthodes hybrides comme l'approche contour-région, pour l'utilisation des champs de Markov et l'algorithme de ligne de partage des eaux couleur issu de la morphologie mathématique.

II.3.1 .3 La segmentation par fusion de régions :

L'approche de fusion de régions est principalement fondée sur l'analyse d'un graphe d'adjacence de régions (GAR) qui analyse une image pré segmentée, constituée d'un ensemble des régions.

Ce graphe d'adjacence de régions (GAR) est une structure de données constituée d'un graphenon-orienté dont chaque nœud représente une région et chaque arête représente une adjacence entre deux régions (Figure II.6).



Figure II.6 Image original (a) et son GAR (b).

Le procédé de fusion consiste à fusionner deux régions adjacentes à condition qu'elles respectent un critère de fusion. Le critère de fusion représente la similarité colorimétrique entre deux régions et la proximité spatiale. Comme pour la croissance de régions, l'ordre du balayage du graphe est important et peut modifier le résultat de la segmentation finale [14].

II.3.1.4 La segmentation par division-fusion de régions :

La segmentation par division-fusion de régions regroupe la segmentation par division et par fusion de régions précédentes. L'image est avant tout divisée suivant une structure de données en régions homogènes qui respectent des critères globaux, puis les régions adjacentes qui répondent à des critères locaux sont fusionnées.

Parmi les structures de données qui permettent d'effectuer cette approche on a : le quadtree et le diagramme de Voronoï [14].

- Le quadtree : est un tétra-arbre dont chaque nœud possède exactement quatre nœuds fils, excepté les nœuds terminaux. Chaque nœud correspond à un bloc, c'est-à-dire une zone de l'image de forme carrée, dont le nombre de liges et de colonnes est une puissance de deux.

Chaque bloc associé à un nœud du quadtree de la partition initiale est analysé de façon récursive afin de décider s'il doit être divisé en quatre sous-blocs. L'analyse récursive s'arrête lorsque chaque sous-bloc respecte un prédicat d'homogénéité. A ce stade de l'analyse, certains blocs adjacents dans l'image sont homogènes selon un prédicat. C'est pourquoi ces couples de blocs sont fusionnés. L'analyse s'arrête lorsqu'il n'existe plus de couple qui respecte le prédicat de fusion.

L'inconvénient majeur du quadtree réside dans la rigidité du découpage carré qu'il impose. Il conduit à un partitionnement global de l'image qui ne respecte pas toujours la forme des régions présentes dans l'image. Par ailleurs, la phase de regroupement des blocs est sensible à l'ordre du parcours du quadtree.

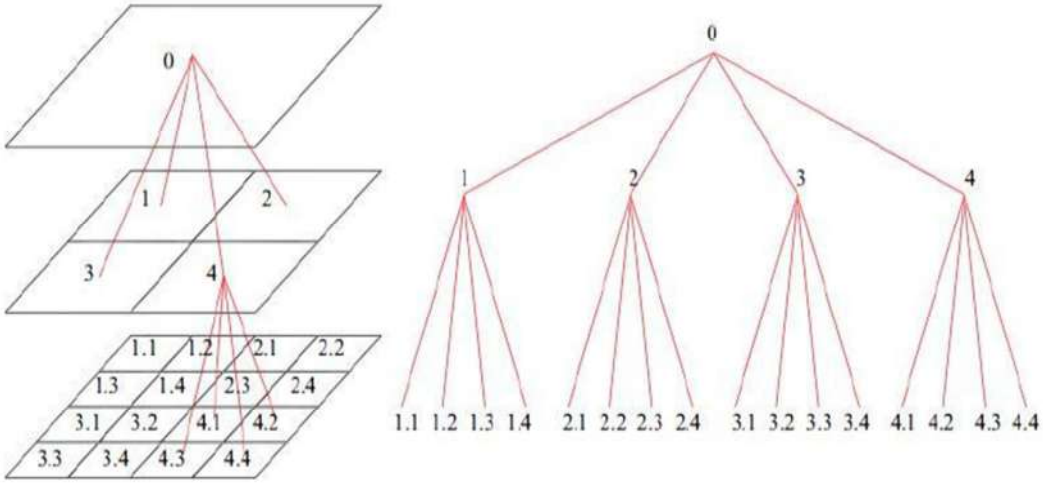


Figure II.7 Structure pyramidale du quadtree.

- Le diagramme de Voronoï :
- le diagramme de Voronoï s'inscrit dans un processus itératif de divisions et de fusions et peut être considéré comme une amélioration de la segmentation par analyse d'un quadtree. La phase de division n'est plus réalisée par un découpage en régions de forme carrée, mais par un découpage en polygones de Voronoï qui s'adaptent aux formes des régions présentes dans l'image. La figure (Fig. 8) représente un exemple de diagramme de Voronoï.

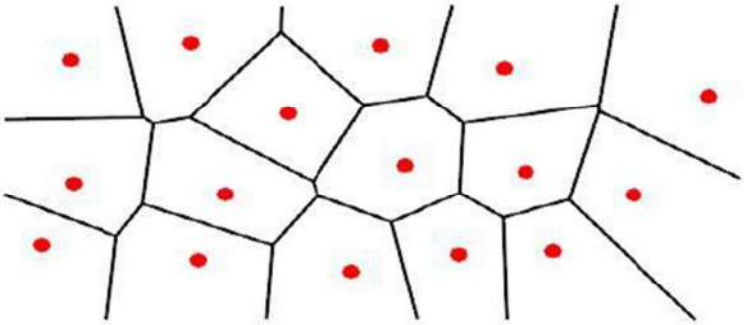


Figure II.8 Exemple de diagramme de Voronoï.

Dans cette figure, les points rouge représentent les germes qu'on positionne initialement grâce à un processus donné (ex : processus de poisson). Les pixels les plus proches d'un germe sont ceux qui sont inscrits dans le polygone noir centré sur le germe [14].

Après avoir présenté les principales méthodes de segmentation d'image couleur par analyse spatiale, nous allons maintenant nous intéresser aux méthodes de segmentation par classification de pixels.

II.3.2.1 Méthodes de segmentation d'images couleur par classification de pixels :

Dans cette section, nous présentons essentiellement les méthodes issues de la classification de données multidimensionnelles qui sont appliquées à la segmentation d'images couleur.

Ces méthodes ne prennent pas en compte la disposition spatiale des pixels et ne considèrent que la distribution des couleurs dans l'espace couleur utilisé. Elles construisent les classes de pixels en présence dans l'image et affectent une étiquette à chaque pixel correspondant à sa classe d'appartenance. La formation des régions n'est obtenue qu'après une analyse de la connexité des pixels dans l'image étiquetée [15].

Comme nous l'avons mentionné dans le premier chapitre, il existe deux types de classifications :

- La classification supervisée : la classification est dite supervisée lorsque des informations à priori sont introduites dans la phase de construction des classes.
- La classification non-supervisée : la classification est dite non-supervisée lorsqu'aucune connaissance à priori n'est disponible.

A l'aide d'une de ces deux types de classification qu'on pourrait construire les classes d'appartenance des pixels. Dans notre travail, on considère qu'aucune connaissance à priori connue sur les images texturées couleur utilisées, donc on s'intéresse aux méthodes de segmentationsd'images couleur par classification de pixels qui sont fondées sur la classification non-supervisées.

Dans les paragraphes qui suivent, on constate que ces méthodes utilisent souvent deux approches [09] pour construire les classes de pixels après avoir représenté les nuages de points qui correspondent aux couleurs des pixels de l'images dans l'espace couleur utilisé (Figure II.9). L'approche métrique et l'approche statistique.

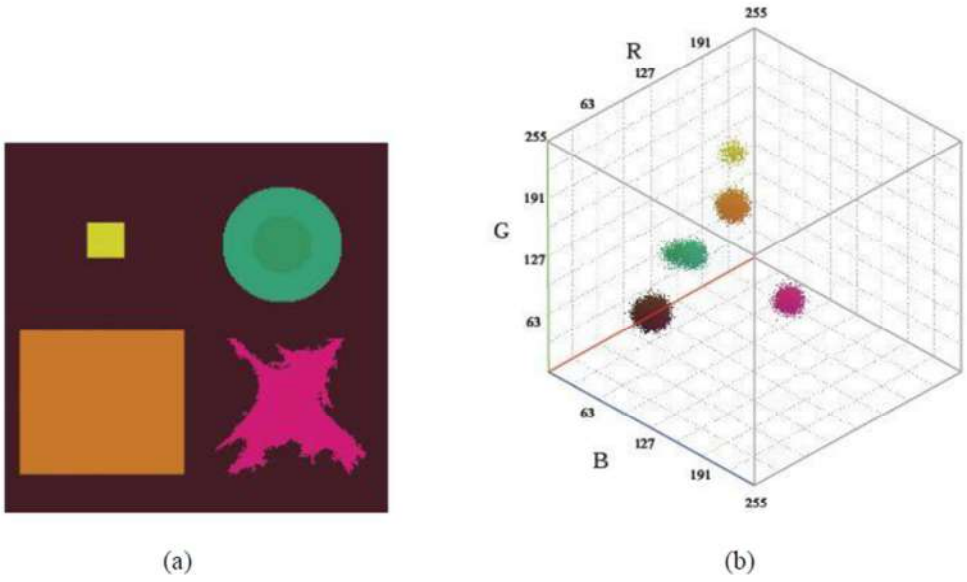


Figure II.9 Nuages de points des pixels de l'image originale (a) dans l'espace utilisé (b).

Certains auteurs proposent de décomposer le problème en l'analyse de plusieurs histogrammes *2D*, (*R, B*), (*R, V*) et (*B, V*) [04] puis chaque histogramme *2D* est analysé en détectant les domaines modaux par une recherche de maxima locaux des contenus des cellules. Une fois le nombre de domaines modaux est connu, ils l'injectent dans un algorithme de fusion de régions qui nous permet finalement d'obtenir l'image segmentée. En résumé, parmi d'autres méthodes qui nous permettent de construire les classes de pixels à base d'histogrammes tridimensionnels, on peut citer l'approche morphologique comme la ligne de partage des eaux (*LPE*) et les filtres morphologiques. Comme les histogrammes tridimensionnels sont très gourmand en espace mémoire, ils sont alors peu utilisés. Pour

remédier à ce problème, les auteurs ont opté pour l'analyse d'histogrammes monodimensionnels qui sont issus de la projection de l'histogramme tridimensionnel sur chacun des axes [14]. Cette solution est avantageuse en espace mémoire mais peut engendrer des erreurs de construction des classes dans la mesure où deux nuages séparables dans l'histogramme *3D* peuvent ne plus l'être sur une projection selon une composante [14].

II.3.2.2 Analyse d'histogrammes monodimensionnels :

De nombreux auteurs déterminent les classes de pixels par analyse récursive d'histogrammes monodimensionnels de la différente composante couleur afin de segmenter les images couleur par des techniques de multi-seuillage [16]. L'analyse récursive des histogrammes monodimensionnels permet de mettre en évidence, au fur et à mesure des itérations, les modes correspondant aux classes de pixels. Elle partitionne l'ensemble des pixels de l'image en différentes classes selon l'algorithme suivant [16]:

- Empiler tous les pixels de l'image dans la pile des pixels à traiter
- Tant que la pile des pixels à traiter est non vide
  - Dépiler les pixels,
  - Calculer les histogrammes monodimensionnels de ces pixels,
  - Rechercher l'histogramme qui présente le mode le plus important,
    - Si le mode est suffisamment important alors
    - Constituer une nouvelle classe avec les pixels qui appartiennent à ce mode,
    - Empiler les pixels n'appartenant pas à ce mode dans la pile des pixels à traiter,
    - Sinon la classe de pixels ne peut plus être séparée en deux. Elle n'est plus empilée dans la pile des classes.
  - Fin Si
- Fin tant que

Les méthodes d'analyse d'histogrammes monodimensionnels proposées dans la littérature se différencient par :

− Les composantes couleur choisies,

− les procédures de détection des modes et de recherche des seuils,

− la détermination de la composante la plus représentative,

− et enfin par les critères d'arrêt de l'analyse récursive des histogrammes

II.3.2.3 Isodata (Iterative Self-Organizing Data Analysis Techniques A):

Isodata est un algorithme recommandé par certains auteurs (K. Takahashi, H. Nakatani et K.Abe) [14] pour la segmentation des images couleur, il est issu de l'approche des k-means qu'on va, voir dans les paragraphes suivants.

Cette méthode permet de partitionner l'ensemble des pixels sans connaitre à priori le nombreexact de classes présentes dans l'image. Elle introduit de nouveaux paramètres afin de modifier lenombre de classes au cours des itérations [14]:

- Lorsque la somme des variances des composantes couleurs des pixels appartenant à uneclasse devient plus grande qu'un seuil fixé, la classe est divisée en deux,
- Lorsque la distance entre les centres de gravité de deux classes devient inférieure à unautre seuil, les classes sont fusionnées.

Elle peut aussi nécessiter d'autres paramètres tels que :

- Le nombre minimal de pixels par classe,
- Le nombre maximum de paires de classes pouvant être fusionnées,
- Le nombre maximum d'itérations.

II.3.3 Méthodes qui nécessitent la connaissance du nombre de classes de pixels :

II.3.3.1 Méthodes des Centre mobiles, nuées dynamiques, k-means :

Les méthodes dites des centres mobiles, nuées dynamiques et des k-means sont très utilisées dans le domaine de la classification de données multidimensionnelles. Elles se ressemblent mais possèdent néanmoins quelques différences qu'on peut citer dans les paragraphes suivants [14]:

Méthode des centres mobiles : elle considère un centre de gravité pour chaque classe. Elle consiste à partitionner l'ensemble des pixels de l'image en un nombre de classes prédéterminé par l'utilisateur. Pour chaque classe, un pixel est sélectionné de façon aléatoire parmi l'ensemble des pixels de l'image afin de constituer son centre de gravité initial. Chaque pixel de l'image est ensuite affecté à la classe dont le centre de gravité est le plus proche. On obtient ainsi une première partition de l'ensemble des pixels.

Une fois que tous les pixels de l'image ont été affectés à leurs classes respectives, le centre de gravité de chaque classe est réactualisé et le processus est réitéré jusqu'à convergence de l'algorithme.

Méthode des nuées dynamiques : cette méthode est une variante des centres mobiles où chaque classe n'est plus représentée par son centre de gravité mais par l'observation la plus proche de ce centre de gravité.

Méthode des k-means : la méthode des k-means diffère de celui des centres mobiles par le fait qu'à chaque itération, chaque nouvelle affectation d'un pixel à une classe entraine la remise à jour immédiate de son centre de gravité. Il est à noter que lorsque deux centres de gravité sont proches, les deux classes associées ne sont pas fusionnées. D'autre part, un centre de gravité isolé dans l'espace des couleurs peut être associé à une classe qui ne contient aucun pixel.

Toutes ces méthodes convergent vers un minimum local de la somme des distances euclidiennes entre les points de chaque classe et son centre de gravité. Ce minimum ne correspond pas nécessairement au minimum global recherché et dépend de la position initiale des centres de gravité.

II.3.3.2 Méthode des Fuzzy c-means :

La méthode des Fuzzy c-means est un algorithme de classification floue fondé sur l'optimisation d'un critère quadratique de classification où chaque classe est représentée par son centre de gravité. Cette méthode nécessite de connaitre le nombre de classes au préalable et génère les classes de telle sorte que les sommes des écarts quadratiques interclasses soient respectivement maximale et minimale [14].

II.4 Conclusion :

Dans ce chapitre, un état de l'art des méthodes de segmentation qui s'adaptent le mieux aux images couleur a été fait (voir les méthodes basées régions). La multitude de méthodes de segmentation qui existe rend le choix de la méthode à utiliser très délicat et notamment dépend aussi de la méthode d'analyse d'image couleur.

On a constaté également dans la littérature que les images couleur peuvent être considérées comme des images multi-spectrales (multi-composantes) et que leurs traitements pouvaient se faire en considérant les différentes stratégies vues précédemment. Pour notre travail, nous avons choisi la stratégie scalaire en raison de sa rapidité et de sa simplicité de mise en œuvre.

Chapitre III

La reconstruction 3D

Chapitre III : La reconstruction 3D

- **INTRODUCTION.**
- **La vision par ordinateur**
- **Description d'un système de vision par ordinateur**
- **La reconstruction 3D**
- **Techniques de reconstruction 3D**
- **CONCLUSION**

III Introduction

La vision est un processus de traitement de l'information. Elle utilise des stratégies bien définies afin d'atteindre ses buts. L'entrée d'un système de vision est constituée par une séquence d'images. Le système lui-même apporte un certain nombre de connaissances qui interviennent à` tous les niveaux. La sortie est une description de l'entrée en termes d'objets et de relations entre ces objets.

La vision par ordinateur est à la base de tout système de vision artificielle. Un système de vision traite les informations de bas niveau d'abstraction du flux vidéo et image pour en extraire des informations de plus haut niveau d'abstraction : présence d'entités, classe d'entités (document, chaise, mur) et plus précisément l'identité de l'entité. Dans ce chapitre, on va donner quelque définition concernant la vision par ordinateur et les systèmes de vision et la classification de ces systèmes puis on parle de la reconstruction 3D en citant quelques méthodes utilisée pour reconstruire des objets 3D.

III.1 La vision par ordinateur

La vision par ordinateur (aussi appelée vision artificielle ou plus récemment vision cognitive) constitue une des branches de l'intelligence artificielle, c'est un domaine d'investigation et d'applications nées à la fin des années 50 dont les premières bases théoriques ont été définies dans les années 60. Depuis, étant donné le spectre très large d'applications, très peu de problèmes ont trouvé des solutions entièrement satisfaisantes. La recherche en vision est divisée en plusieurs approches différentes, chacune étudie des problèmes différents utilisant des techniques différentes.

La vision par ordinateur 2D met l'accent sur les techniques actuelles d'analyse d'image, d'apprentissage et de reconnaissance des formes. La vision par ordinateur 3D approfondit quand a elle les approches permettant d'extraire d'une ou plusieurs images des informations tridimensionnelles relatives à la scène photographiée. Par exemples l'industrielle, militaires, aérospatiales, la télédétection, la sécurité et médical [17].

La vision par ordinateur est la discipline qui cherche à reproduire la perception visuelle humaine sur un ordinateur. Le processus d'acquisition d'images est effectué à l'aide de cameras. Ces images seront traitées afin de produire une description des éléments qui composent la scène 3D.

La figure III.1 explique ce processus qui utilise un ensemble d'images et des connaissances afin d'aboutir à une description d'une scène.

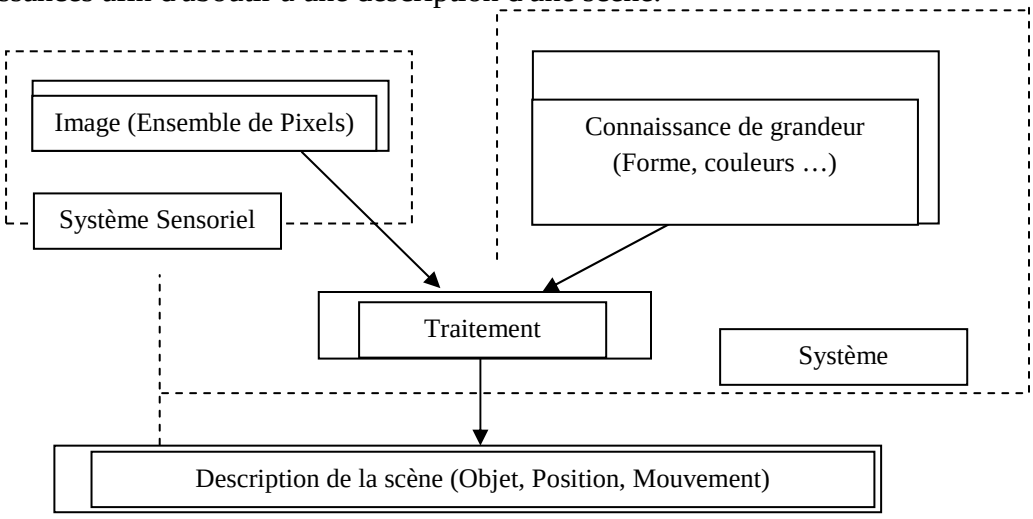


Figure III.1Schéma fonctionnel d'un système informatique pour la vision artificielle.

III.2 Description d'un système de vision par ordinateur

III.2.1 Définition

Un système informatique de vision est un ensemble de processus fortement liés et composés de matériels et de logiciels. A l'aide de ces processus et des images prises de la scène, il fournit en sortie une description symbolique puis sémantique de la scène observée. La description est réalisée à l'aide de l'identification des attributs de la scène et des objets qui la composent.

Le système informatique reçoit en entrée une séquence d'images qu'il traite en utilisant la connaissance spécifique aux différents niveaux de traitements, en sortie il fournira une description en termes d'objets et de relations entre ces derniers.

III.2.2 Classification des systèmes de vision par ordinateur

Selon la nature de la scène observée on détermine le type de système, on distingue deux types de système qui sont :

III.2.2.1 Les systèmes de vision bidimensionnelle

Ces systèmes permettent une acquisition de l'image de la scène 2D, ils n'exploitent pas la notion de profondeur et donc reconnaissent uniquement les objets plans et isoles. Ces

systèmes sont utilisés dans des domaines spatiaux tels que la télédétection et la reconnaissance de caractères, la robotique industrielle...

III.2.2.2 Les systèmes de vision tridimensionnelle

Ces systèmes sont capables d'identifier les objets à partir d'une image d'un monde tridimensionnel, c'est pour cela qu'ils sont utilisés en robotique car ils permettent à un robot de se déplacer dans un environnement inconnu à obstacles.

III.3 La reconstruction 3D

La reconstruction tridimensionnelle d'un ensemble de points informe sur la structure géométrique de ce même ensemble.

III.3.1 Les différentes approches de la reconstruction tridimensionnelle:

- Reconstruction projective

Quand les points sont finis dans un repère projectif. Une telle reconstruction ne contient donc pas d'information critique (absence de la notion d'angle, de longueur, parallélisme ...).



Figure III.2 Une reconstruction projective à droite

- Reconstruction affine

Quand les points sont définis dans un repère affine. Une telle reconstruction contient des informations affines (rapport de longueurs, parallélisme).



Figure III.3Une reconstruction affine à droite

- Reconstruction euclidienne

Quand les points sont définis dans un repère euclidien. Cette dernière reconstruction est la plus riche; elle contient les informations que l'homme a l'habitude de manipuler. Nous parlerons dans ce cas de reconstruction euclidienne, car le facteur d'échelle absent ici ne pose pas de problème particulier.



Figure III.4 Une reconstruction euclidienne à droite

III.4 Techniques de reconstruction 3D

Il existe de nombreuses techniques de reconstruction 3D, On peut répertorier une grande partie des méthodes de reconstruction dans ces quatre catégories :

· **Reconstruction à partir de la texture (Spacecarving) :** Cette méthode basée sur le calcul d'un critère de photo-consistance nécessite d'avoir une lumière constante et homogène tout autour de l'objet.

· **Reconstruction à partir de la géométrie épipolaire :** C'est la méthode la plus classique étant donné qu'on utilise déjà la géométrie épipolaire pour initialiser l'asservissement visuel et que l'on peut estimer les positions des caméras.

· **Reconstruction à partir des contours**: Baser la reconstruction 3D sur les contours de l'objet semble être difficilement envisageable sur un objet naturel de forme quelconque avec des contours très irréguliers.

· **Reconstruction à partir du mouvement de la caméra (Factorisation de Tomasi et Kanade) :** Cette méthode a la particularité de n'utiliser aucun modèle de distorsion. L'extraction précise de points dans les images est également réalisée. Dans un premier temps, la localisation de points dont une modélisation du signal existe est réalisée. Deux types de points sont étudiés : les coins de l'image, et le centre de cibles circulaires. Une méthode de mise en correspondance précise de points entre deux images.

III.4.1 La vision stéréoscopique

III.4.1.1 Principe

La Vision Stéréoscopique vise à reconstruire la structure 3D d'une scène. Afin de réaliser cette reconstruction il est nécessaire de connaitre les coordonnées 3D de tous les points d'une scène. La scène est représentée par un couple d'images stéréoscopique. Ces images sont deux représentations de la scène prises sous des angles différents. Chaque point de la scène est projeté dans les images de la paire stéréoscopique àdes positions différentes. En vision stéréoscopique pour disposer des coordonnées 3D d'un point il faut disposer des coordonnées de ses deux projections, respectivement dans les images gauches et droites. L'obtention des coordonnées de ces deux projections d'un même point 3D est réalisée grâce à une phase d'appariement qui vise à trouver, pour un point donné dans une image, son point correspondant dans l'autre image. Ces deux points sont les projections respectives du même point 3D dans les images droite et gauche.

Le principe de la vision stéréoscopique est résumé dans le schéma synoptique de la figure III.5 il est à noter qu'en amont des étapes de la figure III.5 il est nécessaire de réaliser la calibration du système stéréoscopique de façon à disposer des paramètres intrinsèques et extrinsèques du système. Ces paramètres seront essentiels dans les phases de rectification et de calcul des positions 3D des points.

La stéréoscopie procède par appariement puis une triangulation.

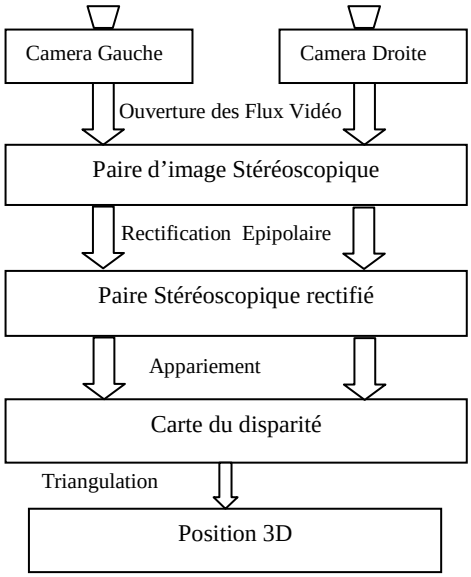


Figure III.5Synoptique général de la vision stéréoscopique.

III.4.1.2 L'appariement

Il consiste à déterminer les projections qui se correspondent dans deux images, c'est à dire quels sont les points 2D représentant les projections d'un même point 3D. Un appariement est donc un ensemble de points 2D en correspondance.

III.4.1.3 La triangulation

C'est une technique d'interpolation de surface, et l'interpolation de surface est le passage du modèle volumique au modèle surfacique. On peut classer la triangulation en deux catégories qui sont :

· Celle qui s'appuie sur une surface définie implicitement.

· Celle qui s'appuie sur un nuage de points représentant la surface de l'objet.

a. La triangulation des surfaces implicites(L'algorithme des marchingcube)

Afin de représenter un volume de données, on divise ce volume en petits cubes, on fixe un seuil de densité par lequel va passer la surface et, à l'aide des points d'intersection entre les cubes et la surface on polygonise cette dernière et on affiche le résultat à l'écran. La première difficulté sous-jacente à cet algorithme est que, si l'on traite tous les cas de façon s'éparée, on a 256 possibilités à traiter (il existe 256 façons qu'une surface courbe peut intersecté les arêtes d'un cube). Heureusement,

En mettant à contribution la symétrie de rotation on peut réduire ce nombre a 23 cas qu'il nous faut traiter s'emparement. On parle ici de 14 cas au lieu de 23 car les symétries de réflexions sont aussi mises à parti. Cependant, après une étude approfondie de la question, il s'est avéré que si les symétries de réflexion sont utilisées des trous ou des inconsistances seront générées dans la surface. Ceci vient du fait que, dans certains cas, la surface se trianguler de façon différente si les sommets du cube qui étaient à l'intérieur de l'objet passent à l'extérieur et vice-versa. Par exemple si l'on regarde les figures III.6 et III.7 qui sont la symétrie l'une de l'autre, on peut voir que la surface peut se trianguler de deux façons différentes (les points noirs représentent les sommets du cube à l'intérieur de l'objet).

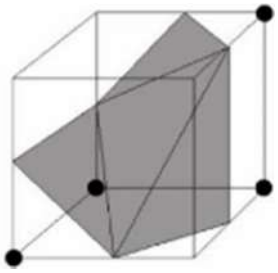


Figure III.6 - Une configuration possible d'un cube triangule.

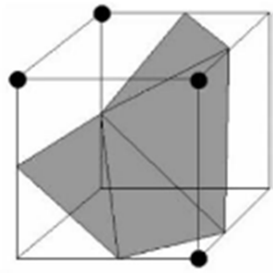


Figure III.7 La configuration miroir d'un cube triangule.

b. Triangulation à partir d'un nuage de points

Construction d'un nuage de points pour chaque **voxel** surfacique en estimant le point par lequel passe la surface, la densité du Vogel et la normal sont utilisé pour estimer la position de ce point.

III.4.2 Reconstruction à partir de géométrie epipolaire

Les principales étapes sont représentées sur la Figure III.8.L'extraction de points et l'appariement se font exactement comme pour l'asservissement visuel, à la différence que l'on travaille sur une séquence d'images. L’étape de calibration de la paire stéréo doit se faire sur

des images prises sur le site sans changer les paramètres intrinsèques et extrinsèques des caméras. Les paramètres intrinsèques varient en fonction du changement de focale. On utilisera une mire car l'auto-calibrage est plus contraignant, nécessitant des mouvements spécifiques de la camera. On supposera que les paramètres intrinsèques ne varient pas durant toute la phase d'acquisition des images. Une première reconstruction projective est alors réalisée pour initialiser la suite de la reconstruction, par une méthode d'ajustement de faisceaux (Bundle Ajustement).

Ensuite un appariement dense des points appartenant à la structure est calculé à partir de, paires d'images rectifiées (La rectification des images est également utilisée pour l'appariement des points servant à l'asservissement des caméras. Enfin une carte de profondeur dense est calculée pour chaque point apparie, puis les points 3D sont reliés entre eux par triangulation (Delaunay). Le plaquage de textures sur les surfaces finit de donner un aspect photo-réaliste à l'objet reconstruit.

Cette méthode de reconstruction doit être appliquée sur chaque sous-ensemble de séquence d'images correspondant à chaque position du robot autour de l'objet, puis les différentes parties reconstruites sont alors assemblées. Il est courant de reconstruire à partir de sous-ensembles d'images puis de fusionner les modèles partiels ainsi obtenus par une phase d'alignement 3D. Les algorithmes de reconstruction et d'alignement s'appuient sur des correspondances de points ou de droites entre les images. La localisation de ces points ou droites dans les images est affectée par un bruit de mesure, influençant la qualité des modèles 3D reconstruits.

Des images 3D sont des images qui représentent une scène en trois dimensions. Le « pixel » est alors appelait voxel, et représente un volume élémentaire.

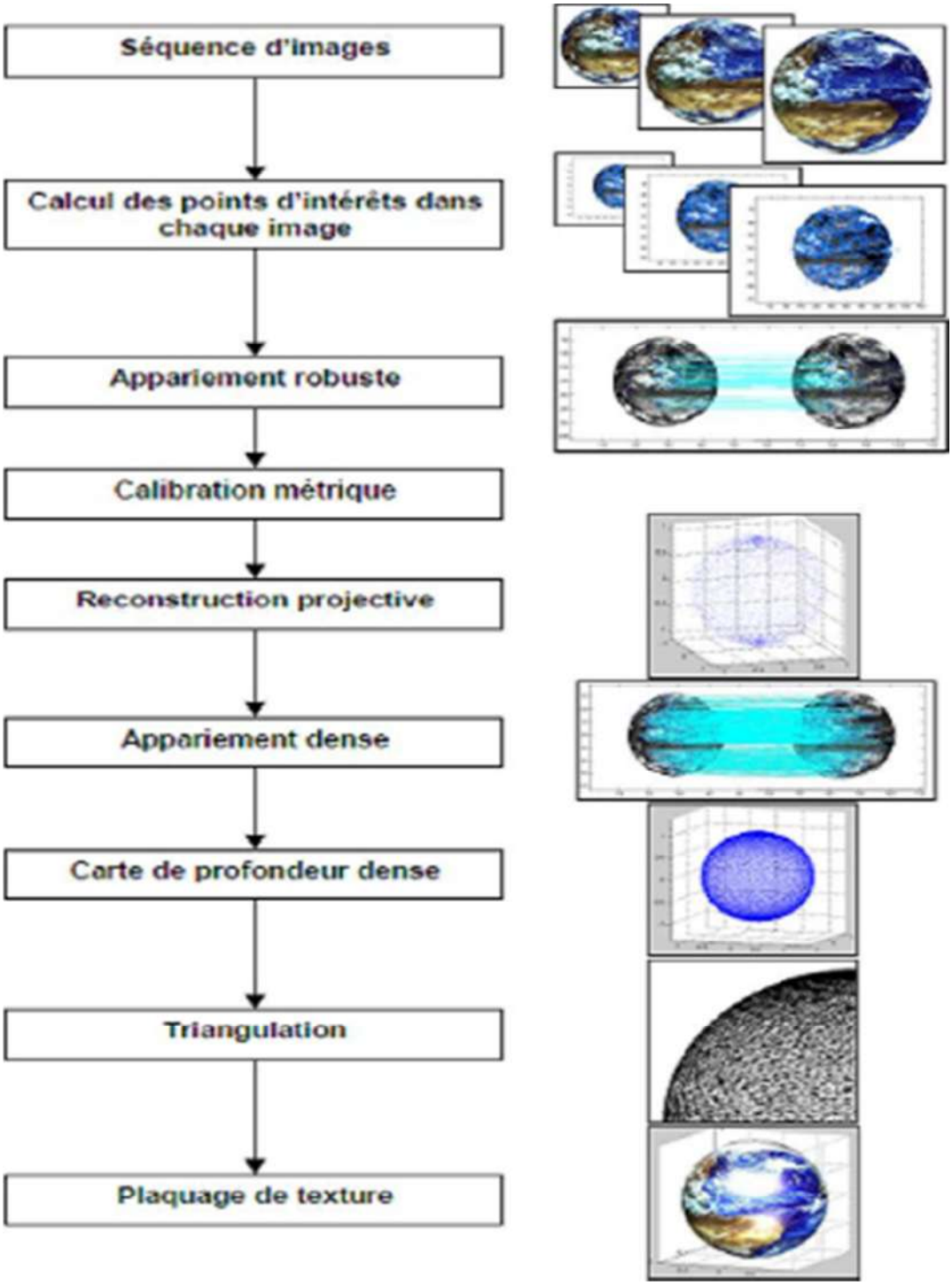


Figure III.8 Méthode de reconstruction.

III.4.3 Technique base sur les ombrages (Shape frome shading)

Comme son nom l'indique, il s'agit d'extraire une forme à partir des variations de lumières observées sur la surface d'un objet [18], [19] au début elle était très peu utilisée dans les systèmes de vision, une des raisons de cet état de fait est l'absence d'algorithmes robustes capables de retrouver les détails fins et discriminants des surfaces des objets (voir figure

III.9). Récemment des études de possibilité d'utiliser le Shape fromshading(SFS) dans la reconnaissance d'objets 3D sont faites.

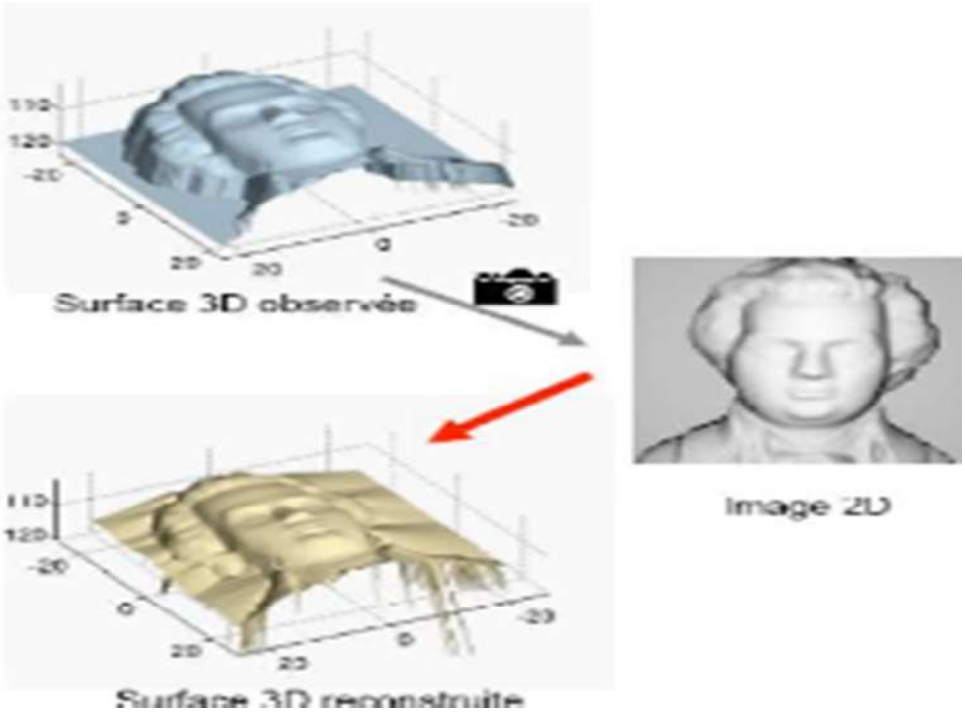


Figure III.9 Application de la SFS sur une image 2D.

III.4.4 Représentation par détection de déformation de formes

Cette approche décrit une nouvelle représentation des formes déformables basées sur des polygones triangulaires (voir figure III.10), [20][21][22]elle cherche une structure de correspondance de calibre déformable, ou on veut trouver une transformation non rigide qui dresse la carte d'un modèle à une image. Le problème posé est la correspondance de la forme à l'image, ce dernier est défini en termes d'une fonction d'énergie. Cette fonction d'énergie associe un cout à chaque transformation potentielle du modèle. Typiquement la fonction d'énergie est une somme de deux termes, le premier rapproche le modèle déformé, des caractéristiques d'image, tandis que le deuxième pénalise les grandes déformations du modèle. La recherche sur des transformations est faite efficacement en exploitant les propriétés de cette représentation pour les formes déformables. Cette approche donne une solution optimale globale au problème de correspondance qui permet de détecter des objets déformables sans aucune sorte d'occlusion et le désordre de fond.



Figure III.10 Polygones triangulaires d'un lapin et d'une poire.

III.4.5 Méthodes basées sur l'extraction de squelette

La représentation des formes en squelette est introduite par Blum [23]. Elle a la possibilité de conserver sous forme compacte un grand nombre d'informations topologiques et géométriques de la forme initiale. Un autre avantage qu'il faut signaler est le fait que les squelettes ont une structure de graphe (voire figure III.11) qui permet d'utiliser les puissants outils issus de la théorie des graphes. Le squelette d'un objet est une représentation de la forme dans une dimension inférieure. Pour un objet surfacique (2D), le squelette est un ensemble de lignes centrées dans la forme, par contre pour les objets 3D il existe deux types de squelettes. Les squelettes surfaciques qui sont un ensemble de surfaces centrées dans la forme et les squelettes curvilignes 3D qui sont un ensemble de lignes centrées dans la forme.



Figure III.11 Squelettes topologiques de quelques objets.

III.4.6 La reconstruction par des jonctions caractéristiques

Cette méthode est proposée par R.BERGEVIN et A.BUBEL [24]. Elle est basée sur l'extraction des branches de jonction d'objets tridimensionnels aux points d'intérêt dans une image 2D en utilisant un processus de regroupement topologique.

Pour la reconstruction des objets 2D, avec cette méthode il faut construire un graphe, ou les nœuds sont les points de jonctions, les arcs sont les contours participant à la jonction, en suite ce graphe est exploité pour extraire la structure des contours composant l'objet .Le principe de cette méthode est de subdiviser l'image en petites fenêtres pour faciliter l'extraction des jonctions possible. Chaque fenêtre est analysée en la binarisant, puis on extrait les branches en appliquant des opérateurs qui mettent en évidence les lignes de direction prédominantes. A la fin, l'ensemble des branches extraites est organisées en graphe et les positions de jonction extraite sont structurellement validées puis raffinés.

III.4.7 La reconstruction à partir du mouvement (Shape From Motion)

L'analyse du mouvement et l'interprétation visuelle d'une scène sont extrêmement liés. En particulier, le mouvement du capteur autour d'un objet peut permettre à l'observateur de mieux comprendre la forme et la structure de l'objet. En réalité l'origine du mouvement revient à trois facteurs:

· Le mouvement de l'observateur en respectant la scène (Déplacement de la caméra dans le cas de la stéréoscopie): Ou le mouvement des objets dans la scène :La camera et l'objet de la scène se déplacent.

Le domaine de suivi de mouvement se divise en deux tâches principales :

· Détection de mouvement : consiste à trouver les objets en mouvement dans la scène par l'utilisation des séquences d'images captées à des moments différents: Estimation de mouvement : consiste à estimer le déplacement de certaines primitives de l'image.

Différentes approches (approche de flot optique, approches de mise en correspondance; sont utilisées pour la détermination de mouvement et la structure tridimensionnelle de l'objet. Cette dernière, consiste à trouver les coordonnées 3D d'un point P à partir d'une séquence d'images. En déplaçant une caméra autour d'un objet, on peut construire sa structure tridimensionnelle.

III.4.7.1 L'approche basée sur la mise en correspondance

Cette approche utilise les techniques de l'appariement des indices pour l'estimation du mouvement et le suivi d'objets, ces primitives sont généralement des coins, des segments. La suite des positions de ces indices visuels dans plusieurs images permet, si l'on utilise suffisamment d'indices appartenant au même objet d'obtenir le mouvement 3D de celui-ci. Le principe consiste en un système à huit équations de la forme :

$$E = T \times R$$

Tel que T est une translation et R est une rotation, connaissant huit points de l'objet, la décomposition de E en valeurs singulières permet de retrouver les paramètres du mouvement et la structure 3D. L'estimation du mouvement à partir des droites donne de meilleurs résultats que ceux donnés par des points caractéristiques.

III.4.7.2 L'approche du flot optique

L'utilisation de cette méthode consiste à extraire l'information de la vitesse à partir d'une séquence d'images avec l'hypothèse que l'intensité des pixels des objets est conservée au cours du déplacement. Le calcul du flux optique consiste à extraire un champ de vitesses à partir d'une séquence des images en supposant que l'intensité(ou la couleur) est conservé au cours du déplacement. Sous cette hypothèse, on peut établir une relation entre la vitesse apparente V (déplacement dans l'image d'indices visuels tels que des régions délimitées par des contours supposé représenter la projection du mouvement 3D des objets de la scène et /ou du mouvement de la caméra) et les variations spatio-temporelles de l'intensité.

Le principe de cette technique est résumé dans les quatre étapes suivantes :

1. La caméra prend plusieurs images de la scène.

2. Détection de mouvement (approche de flot optique, mise en correspondance).

3. Estimation de mouvement (angles de rotation, matrice fondamentale, vecteurs de translation . . . etc.)

4. Estimation de la carte de profondeur.

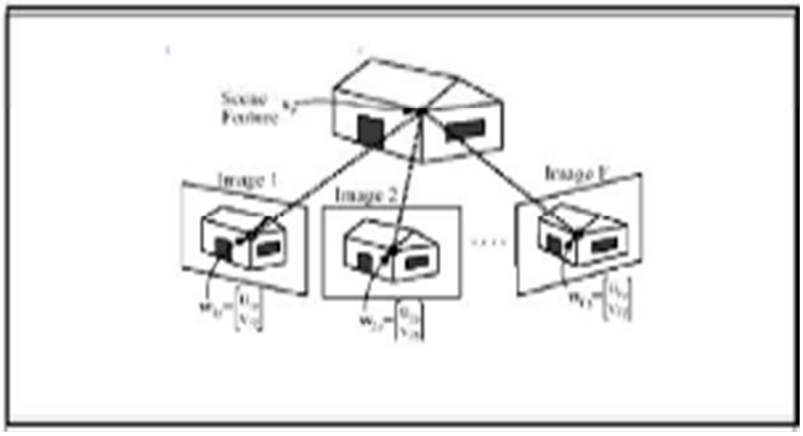


Figure III.12 Processus de reconstruction 3D à partir d'un mouvement

Reconstruire le modèle 3D d'après la carte de profondeur calculée.

III.4.8 Méthode par space carving (creusage d'un volume 3D)

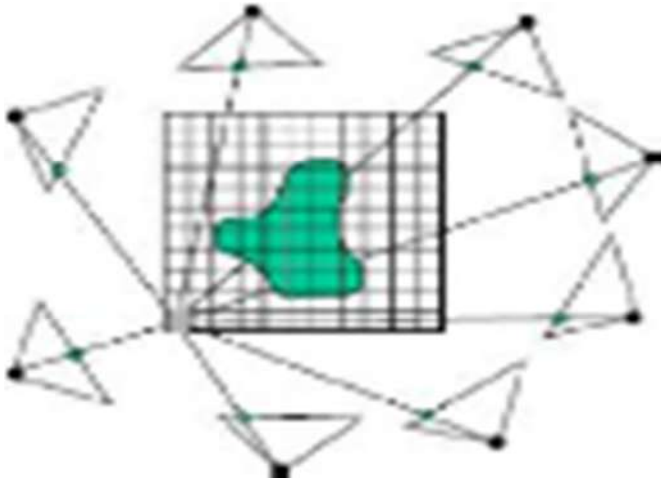


Figure III.13 - Méthode par space de l'objet

· Algorithme de spacecarving [26])

1. Initialiser un volume englobant fait de N x M x P voxels

2. Choisir un voxel sur la surface courante

3. Projeter ce voxel sur les caméras sur lesquelles il est visible

4. Éliminer (ou << creuser >>) si non photo-cohérent

5. Répéter 2-3-4 jusqu'à convergence

III.5 Conclusion

Dans le présent chapitre on a abordé la notion de la vision par ordinateur et la définition correspondante au système de vision par ordinateur et la classification de ces systèmes selon un critère de profondeur.

On a aussi introduit un aperçu sur quelque méthode de reconstruction 3D en montrant leurs définitions et leurs principes.

Chapitre IV
Application et résultats

Chapitre IV : Application et résultats

- **INTRODUCTION.**
- **La Kinect**
- **Principe de Fonctionnement.**
- **Description de la méthode utilisée pour notre application**
- **L'algorithme Résultats et Discutions.**
- **CONCLUSION**

IV. Introduction :

Ce chapitre va être le fruit de tous ce qui a été fait au paravent. Nous allons présenter les résultats d'implémentation de la méthode de segmentation RGBD par la méthode de détection des objets par le billé d'histogramme et d'autre part de présenter les résultats de cette méthode en utilisant des méthodes d'évaluation.

L'implémentation des algorithmes a été réalisée en utilisant le logiciel MATLAB version (R2013b).

IV.1 La Kinect :

Kinect, est un périphérique destiné à la console de jeux vidéo Xbox 360, permettant de contrôler des jeux vidéo sans utiliser de manette. Le nom « Kinect » est issu des mots anglais « *kinetic* »traduiten « cinétique »et « *connect* » veut dire « connecter ».

Il s'agit d'une caméra utilisant des techniques d'interaction développées par la société israélienne PrimeSense, longtemps nommée par le nom de code du projet, « Project Natal », elle a été officialisée juste avant l'E3[1] sous le nom Kinect. Elle est basée sur un périphérique d'entrée branché sur la console Xbox 360 qui permet d'interagir par commande vocale, reconnaissance de mouvement et d'image.

On peut ainsi jouer sur des jeux spécialement développés pour le projet, donc sans aucune manette ni périphérique autre que son propre corps. Cette détection de mouvements sans périphériques était déjà présente dans l'EyeToy de Sony. Cette technologie est destinée au grand public avec une expérience proche des party games de la Wii[2]. La technique a été annoncée lors du salon de l'E3 2009.

Le 21 octobre 2010, Microsoft dévoile la première publicité officielle de son nouveau périphérique, montrant à quel point la firme vise le grand publique.

1 : Abréviation de Electronic Entertainment Expo. Salon regroupant chaque année, à Los Angeles, les principaux acteurs du marché du jeu vidéo.
2 : C'est une console de jeu du fabricant Japonais Nintendo

IV.2 Principe de Fonctionnement :

IV.2.1 Concept :

Ce périphérique de Microsoft est constitué d'une barre horizontale connectée à sa base via un petit moteur. Celui-ci permet à la caméra d'effectuer des petits mouvements vers le haut ou le bas (27° semble-t-il) afin d'adapter la perception de la caméra en fonction de votre position dans la pièce. La barre horizontale constitue l'élément principal de la technologie Kinect. Elle contient une série de multi-microphones, une caméra RGB, et enfin un « 3D depthsensor » permettant d'améliorer l'analyse de mouvement par rapport à une caméra comme l'EyeToy. L'association de ces trois technologies devrait ainsi permettre la reconnaissance vocale et faciale de l'utilisateur, ainsi que la « capture » de l'ensemble de la pièce (et donc du corps) en 3D.

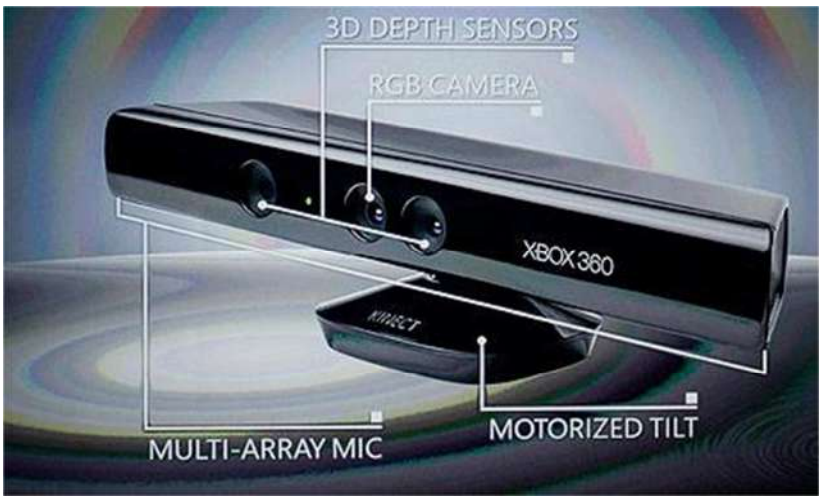


Figure IV.1 Un petit récapitulatif des technologies embarquées dans Kinect.

On va donc commencer cette autopsie par une courte présentation des microphones. Kinect embarquera vraisemblablement 2 microphones ainsi que 4 détecteurs digitaux externes de sources audio. La combinaison de l'ensemble de ce système audio permet ainsi à Kinect de détecter la localisation spatiale d'une source sonore mais aussi d'éliminer les bruits de fond parasites grâce à un traitement de données efficace.

Un microphone va fonctionner de la même façon qu'une oreille humaine : celui-ci va traduire l'onde sonore, émise par une source, en signal électrique. Lorsque l'onde sonore vient toucher / exciter la membrane (aussi appelé diaphragme) du microphone, l'excitation de cette membrane va se traduire en signal électrique analogique avant d'être convertie en signal

numérique. Le diaphragme est l'équivalent du tympan dans notre oreille et son fonctionnement est identique en tout point. Pour convertir un signal analogique en signal numérique, il est obligatoire d'utiliser un convertisseur analogique/numérique. Celui-ci est généralement constitué d'un montage électronique dont la fonction est de traiter une valeur analogique, pour ensuite générer une valeur numérique. Cette valeur numérique sera proportionnelle à la valeur analogique permettant de coder sur plusieurs niveaux (bits).

IV.2.2 La camera RGB (Red Green Blue):

La première des deux caméras embarquées dans la technologie Kinect est une caméra couleur RGB « standard » avec un capteur photographique de type CMOS. Elle se situe au centre de la barre horizontale (Figure IV.1). Elle permet une prise d'image avec une fréquence de 30Hz, en couleur 32bits et en résolution VGA de 640x480 pixels.

IV.2.2.1 Le capteur photographique :

C'est un composant électronique sensible à la lumière qui va convertir un rayonnement (Ultra-Violet, Lumière visible ou Infra Rouge) en un signal analogique (une nouvelle fois via un signal électrique). Ce signal est ensuite numérisé par un convertisseur analogique-numérique afin d'obtenir une image numérique. Les rayonnements et la lumière en quelques termes sont : des ondes électromagnétiques **v**isibles par notre œil. Ces ondes possèdent une « longueur d'onde » (une fréquence) propre pour chaque couleur que nous connaissons. Lorsque cette longueur d'onde est comprise entre environ 400 et 800 nanomètres, nous sommes dans la gamme des couleurs et de la « lumière visible ».

Pour des longueurs d'ondes inférieures ou supérieures, nous retrouvons les rayonnements avec les domaines des ultraviolets, des infrarouges, des rayons X ou des ondes radio. (Figure IV.2)

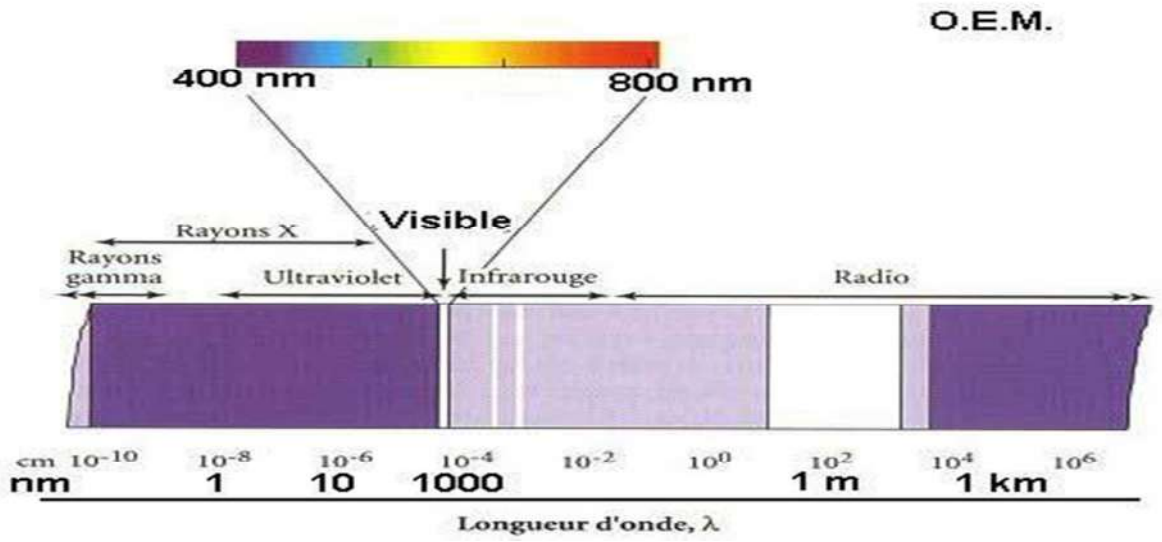


Figure IV.2 : Représentation du spectre électromagnétique.

La lumière est composée de 3 couleurs primaires qui sont le Rouge, le Vert et le Bleu comme illustrées sur la Figure IV.3. L'association de ces 3 couleurs permet d'obtenir la lumière blanche. Le noir représente l'absence totale de couleur. Chaque couleur est donc définie selon le système RVB (ou RGB en anglais) c'est-à-dire en fonction de la proportion respective en rouge, vert et bleu. Le mélange de ces trois couleurs compose aussi chacun des pixels de nos écrans LCD, LED...

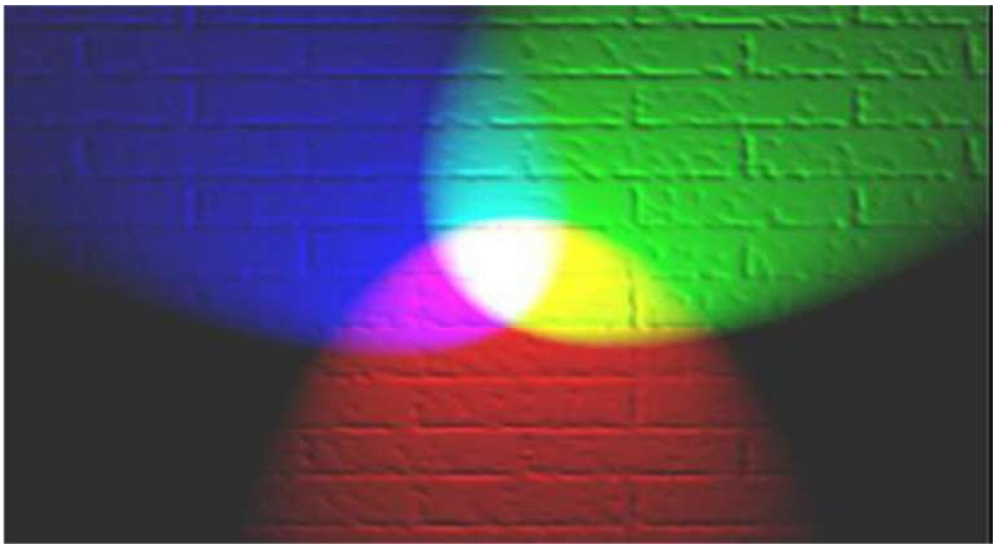


Figure IV.3 Contrairement aux couleurs primaires en peinture (cyan, magenta et jaune).

Pour convertir le signal analogique reçu en image numérique, le capteur photographique va filtrer la lumière selon ces 3 couleurs et sortir trois signaux numériques

correspondant à chacune d'elles. Actuellement, deux grandes familles de capteurs sont disponibles : les CCD et les CMOS.

Les CCD sont surtout utilisés dans les appareils compacts mais sont de plus en plus délaissés dans les reflex. Le principe des photos capteurs CCD (demandant des connaissances en physique-chimie) car le résultat final est le même que pour la technologie CMOS.

Le CMOS est représenté sous la forme d'un petit « écran » de taille variable comme illustrer sur la photographie ci-dessous. La partie sensible à la lumière est la partie grisâtre au centre. Certains capteurs CMOS peuvent actuellement atteindre une résolution hallucinante de 25 millions de pixels et la technologie ne cesse de progresser.

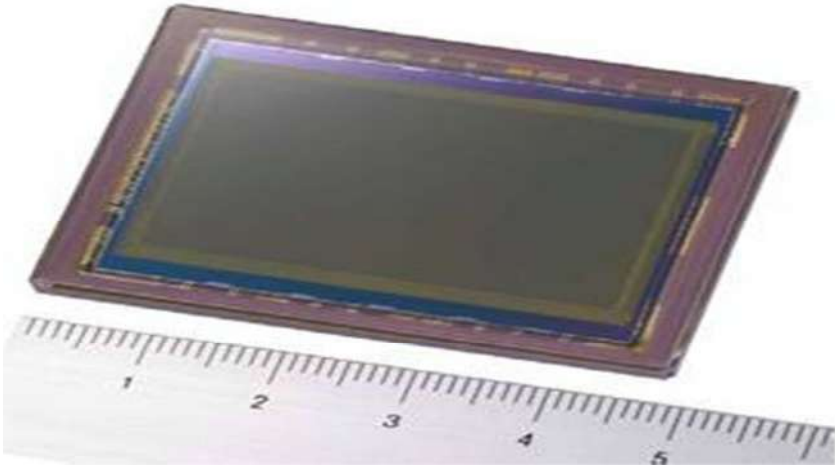


Figure IV.4 : Capteur photographique de type CMOS développé par Sony.

IV.2.2.2 Fonctionnement du capteur :

La lumière (l'image) arrive en face du capteur, celle-ci est d'abord purifiée par un filtre Infra Rouge (bloquant les ondes infrarouges et laissant passer la couleur), puis traverse un « mini filtre » de couleur rouge, vert ou bleu placé en face du capteur lui même. Ainsi les millions de capteurs présents à la surface du CMOS vont émettre un signal électrique relatif à une des 3 couleurs, qui sera ensuite converti numériquement.

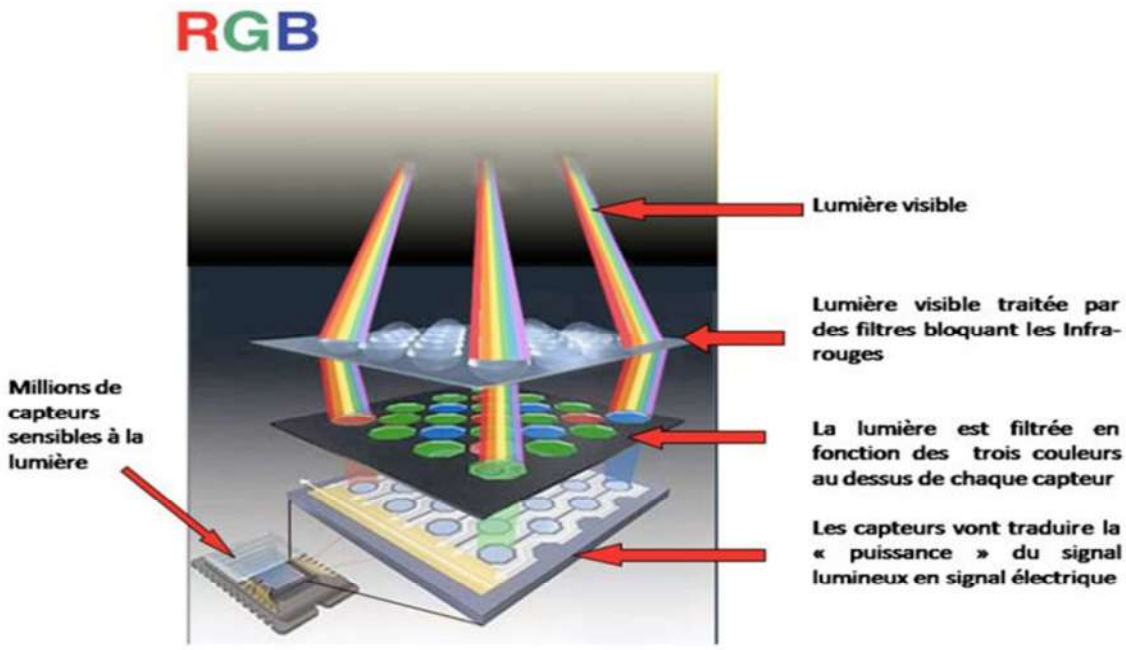


Figure IV.5 : Fonctionnement d'un capteur CMOS pour les appareils photos numériques..

Les propriétés de l'appareil photo ou de la caméra vont dépendre majoritairement du capteur CMOS (ou CCD) présent dans l'appareil. Ainsi pour les téléphones portables, des capteurs de plus en plus petits sont fabriqués afin d'économiser le maximum de place, au détriment de la qualité de l'image obtenue.

Après avoir vu le principe du capteur CMOS pour la caméra RGB, nous allons pouvoir passer aux choses sérieuses avec le « 3D depthsensor ».

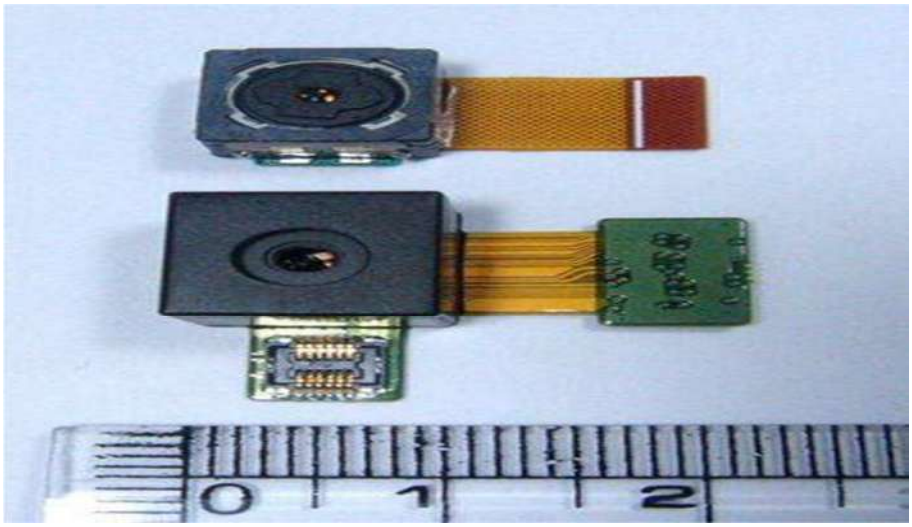


Figure IV.6 : Capteur photographique de taille réduite pour les téléphones mobiles.

IV.2.3 Le 3D Depth Sensor

Jusqu'à présent nous avons vu les technologies microphone et caméra RGB « classique » mais c'est véritablement avec sa technologie « 3D depthsensor » (comprenez capteur de profondeur/distance 3D) que Kinect tire son épingle du jeu. Cette technologie a été proposée par la société israélienne PrimeSense. La caméra de droite sur la figure IV.7 est une caméra infrarouge (pas de couleur) à capteur CMOS, QVGA de résolution 320x240, 16bits (donc possédant 65536 niveaux de sensibilité). L'objectif à gauche de Kinect n'est pas une caméra, comme beaucoup pourraient le penser, mais un émetteur de lumière infrarouge. C'est pourquoi certaines images de Kinect sont représentées avec une lampe lumineuse comme sur la figure IV.7.



Figure IV.7 : Illustration des émissions infrarouges non visible par notre œil par la Kinect

IV.2.3.1Comment fonctionne un capteur CMOS infrarouge :

De façon identique à une caméra RGB sauf qu'on laisse passer uniquement les infrarouges (revoir figure IV.2). La caméra infrarouge permet d'obtenir une image représentant les dégagements thermiques émis par l'objet observé. Ce type d'images furent largement popularisées par le film *Prédateur* et sa mythique vision infrarouge. Cependant, les images obtenues ne sont pas colorées, ce sont les utilisateurs qui décident de rajouter les niveaux de couleurs en fonction de la température mesurée.

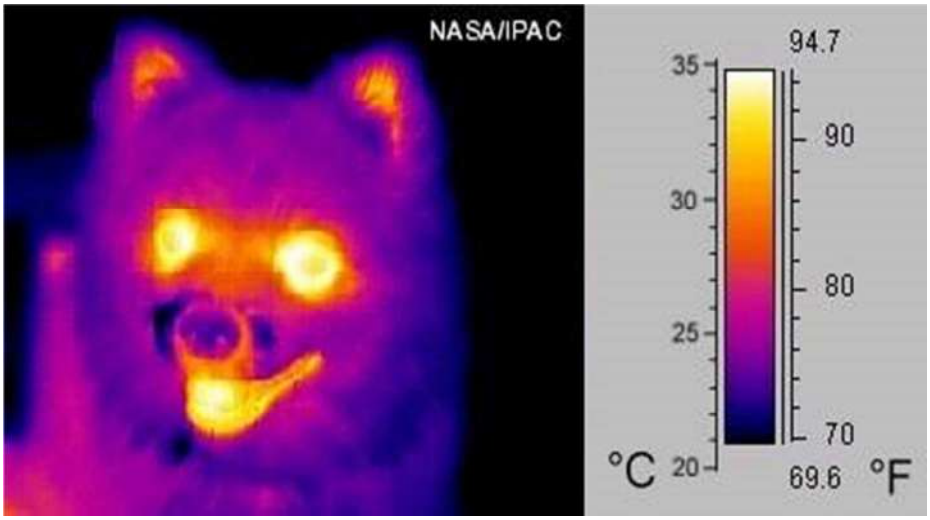


Figure IV.8 : Emissions de chaleurs qui proviennent essentiellement des yeux et de la bouche.

Mais, peut-on faire un détecteur de distance avec une image thermique ? Les yeux du chien sont bien plus éloignés que le bout de sa truffe ! ». Et bien tout simplement, parce que, Kinect ne s'appuie pas sur la chaleur émise par notre corps. En effet, tout l'intérêt de Kinect provient de son émetteur (lampe) de lumière infrarouge**.**

La scène contenant les joueurs est bombardée par les rayons infrarouges non visibles à l'œil. Une partie de ces rayonnements va être réfléchie par l'ensemble des surfaces touchées. Plus l'objet sera loin et plus la quantité de rayonnement infrarouge réfléchie (renvoyée vers la caméra) sera faible. A l'inverse, plus l'objet sera proche et plus la quantité de rayonnement infrarouge réfléchie sera importante. Ainsi la caméra infrarouge va mesurer la distance de l'objet en fonction de l'intensité comme illustré sur l'exemple ci-dessous.

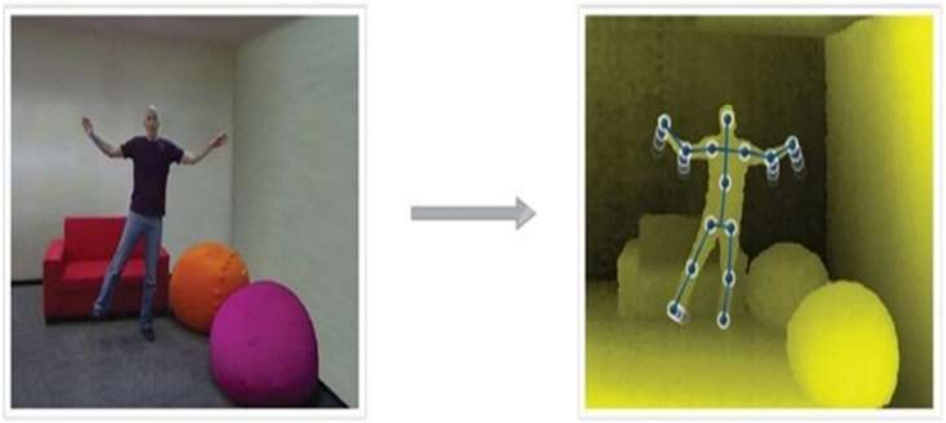


Figure IV.9 : Une scène réelle et l'image recueillie par la camera Kinect.

Le pouf à l'avant est de couleur jaune symbolisant un objet plus proche. S'il était observé par une caméra infrarouge classique, il serait invisible avec le fond du décor car il n'émet aucune

chaleur. Grâce à la différence de couleur entre le joueur et le fond, il est facile de distinguer l'ensemble de sa silhouette, mis à part les pieds en contact avec le sol. Ensuite, ce sont les logiciels qui vont effectuer le traitement des images reçues, en intégrant les différentes parties du corps à l'image.

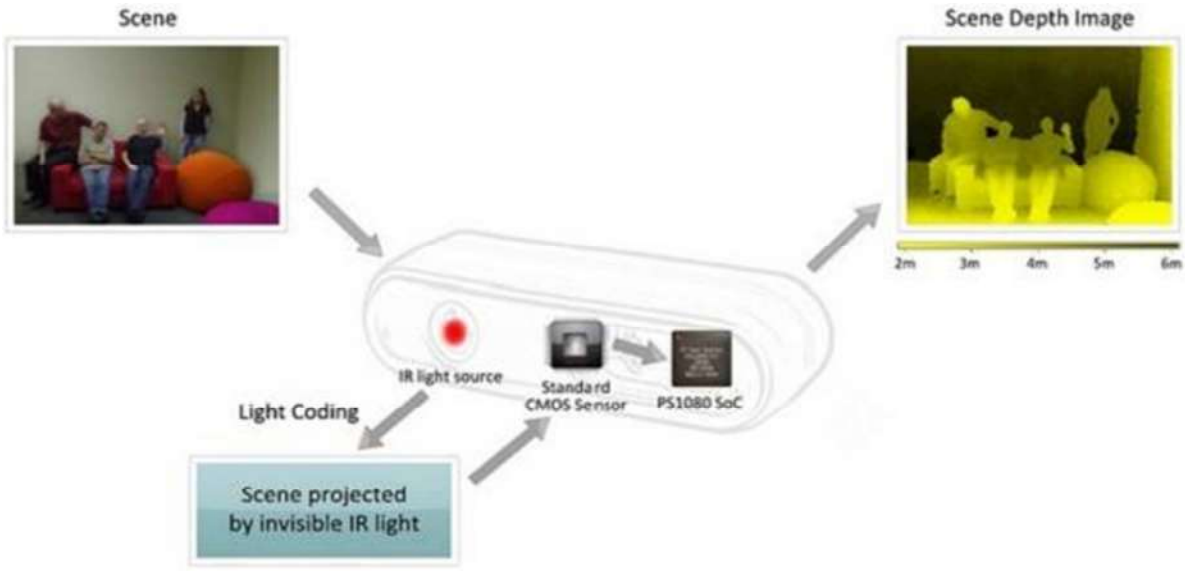


Figure IV.10 : Schéma résume la technologie Kinect.

Ainsi, d'après la figure IV.10 (image de droite), il est possible de cartographier avec précision la distance pour tout objet éloigné de 1,5 à 2 mètres de la caméra jusqu'à environ 4-5 mètres de profondeur. Au-delà de 5 mètres, le rayonnement IR réfléchi devient trop faible pour être mesurable avec précision. Pour tout objet dont la distance est inférieure à 2 mètres, le phénomène inverse est observé et le signal devient totalement saturé**.** La publication récente des distances recommandées confirme cette limitation technique : la distance pour pouvoir jouer à deux joueurs est de 2,4 mètres.

L'avantage de l'utilisation de la lampe IR est de pouvoir jouer dans toutes les conditions de luminosité ! Même si certaines fonctionnalités devraient être affectées comme la reconnaissance faciale ou le scan des objets (utilisant la caméra RGB, dépendante du spectre de lumière visible).

IV.3Description de la méthode utilisée pour notre application :

IV.3.1 Le Matériel utilisé :

IV.3.1.1 A propos ASUS Xtion PRO LIVE RGB and DepthSensor :

La solution de développement Live Pro Xtion permet aux développeurs d'appliquer la dernière technologie de détection de mouvement dans diverses applications, industries, le marketing, l'orientation ou le tourisme et ainsi de suite. La réadaptation physique est aussi un bon terrain pour les applications de détection de corps. Il ya un certain nombre de domaines qui pourraient bénéficier de la technologie basée sur le mouvement, tels que l'éducation, médical, des conférences, des jeux et beaucoup plus.

Le kit de développement LIVE PRO Xtion est largement ouvert. Vous pouvez créer vos propres applications pour les affaires ou pour rendre la vie des gens plus facile et intuitive. N'importe qui peut développer leurs propres applications. Une application de grande envergure peut être créé et appliqué sur la base de l'imagination des développeurs



Figure IV.11 Kinect ASUS Xtion PRO LIVE RGB and Depth Sensor

IV.3.1.2 Principe :

Basé sur le même principe de fonctionnement que la Kinect XBOX 360 déjà expliqué. Le Xtion Pro Live utilise des capteurs infrarouges, la technologie de détection de profondeur adaptatif, image couleur détection et flux audio pour capturer en temps réel l'image, le mouvement et la voix des usagers, ce qui rend le suivi des utilisateurs plus précis. Le PRO solution de développement LIVE Xtion est livré avec un ensemble d'outils de développement

pour le rendre plus facile pour les développeurs de créer leurs propres applications gestuelles sans la nécessité d'écrire des algorithmes de programmation complexes.



Figure IV.12Kinect ASUS Xtion PRO LIVE RGB

a. **Détection de Geste:** Le PRO solution de développement LIVE Xtion écoute les mouvements de la main des gens sans aucun retard, qui transforme votre main dans un contrôleur. Il a plus de 8 poses prédéfinies pour vous permettre de pousser, clic, cercle, vague et beaucoup plus parfait pour une utilisation, commandé une interface utilisateur. Basé sur cette fonction, il peut être largement développé dans divers types d'applications, un gain de temps et d'efforts pour le développement de logiciels de développeur.

b. **Tout détection de corps:** Le PRO solution de développement LIVE Xtion permet aux développeurs de suivre tout mouvement du corps des usagers, ce qui le rend idéal pour toute jeu de corps, tout en soutenant la reconnaissance de lecteur multiples.

c. **Camera RGB**: Xtion Pro Live permet la détection des couleurs (RVB) d'image. Xtion PRO LIVE peut capturer l'image des utilisateurs, ce qui est utile pour la détection humaine, la signalisation numérique, système de sécurité et d'autres applications à être créés.

d. **Audio:**le flux audio permet de soutenir la commande vocale et d'autres applications de reconnaissance vocale. La vidéoconférence au bureau et à domicile sont également possibles avec la fonction RVB.

e. **Facile à utiliser :**Le Xtion Pro Live dispose d'une prise facile et jouer conception USB, avec le Xtion PRO développement, LIVE solution OPNI NITE middleware compatible. S'il vous plaît visitez le site Web OpenNI plus de soutien technique et de l'information.

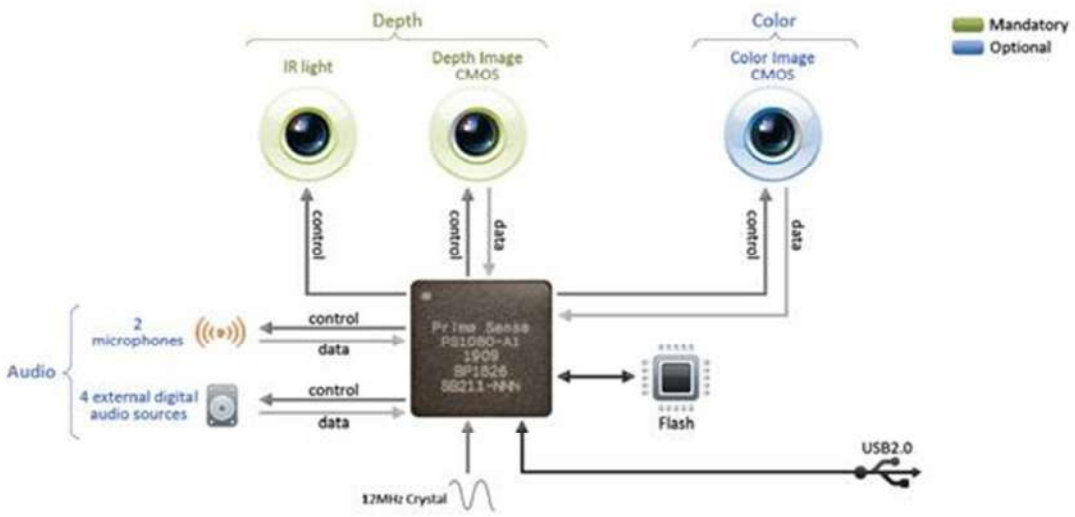
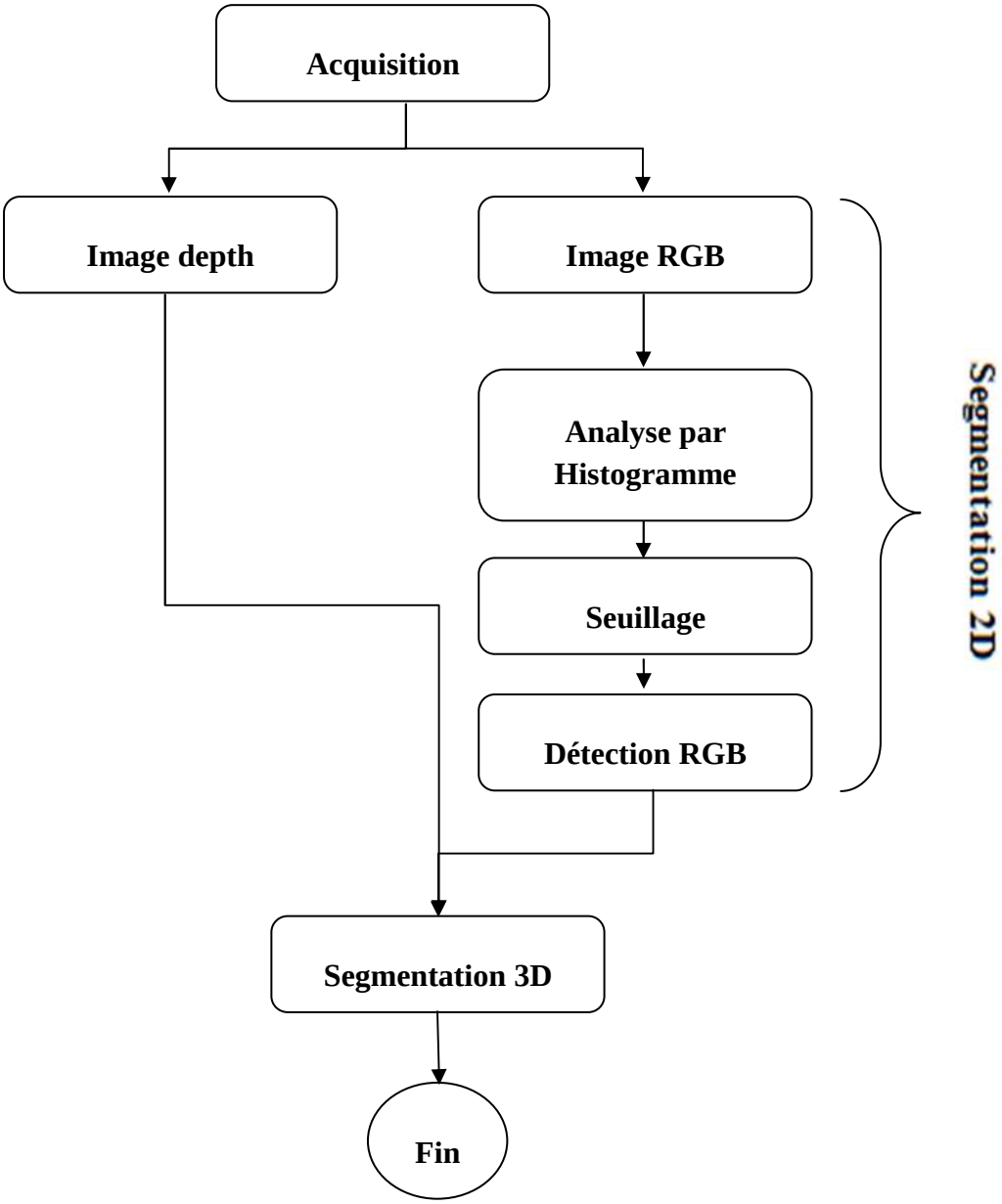


Figure IV.13 Schéma résumant l'ensemble des technologies Kinect

IV.4.L'algorithme :



IV.5Résultats et Discutions:

IV.5.1 Les Expériences

Expérience 1la segmentation suivant l'axe Z: dans le premier cas on a positionné les objets dans l'espace de telle façon qu'ils soient éloignés de différant distance par rapport à l'axe Z de la Kinect.



Figure IV.14 Image Original 1

Les statistique métriques réel de positionnement des objet (rouge,vert, bleu) entre eux, ainsi la kinect sont représenter par le tableau suivant:

Tableau IV.1 Distances Des objets Prisent Manuellement (Exp1)

	rouge			vert			Bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
Kinect	900	0	2500	0	0	2500	1000	0	2500
bleu	1980	50	0	980	100	0			
vert	1000	100	0				980	100	0
rouge				1000	100	0	1980	50	0

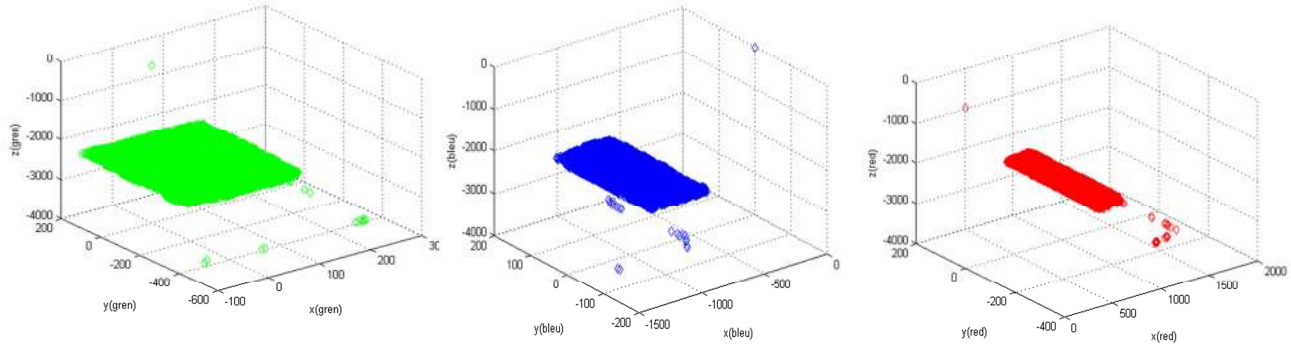
Apres le traitement on obtient les résultats de la segmentation 2D à base d'histogramme:



a **b** **c**

Figure IV.15 Segmentation 2D:a)segmentation des objets rouges, b) segmentation des Objets vert, c)segmentation des objets bleu.

Dès que la segmentation 2D est faite on la combine avec l'image depth pour obtenir la segmentation 3D les résultats de la segmentation en 3D est représenter par la figure suivante:



a **b** **c**

Figure IV.16Segmentation 3D :a)segmentation des objets rouge l'espace, b) segmentation des objets vert dans l'espace, c)segmentation des objets bleu dans l'espace.

Apres la Reconstruction 3D de tous les l'objet on obtient le résultat:

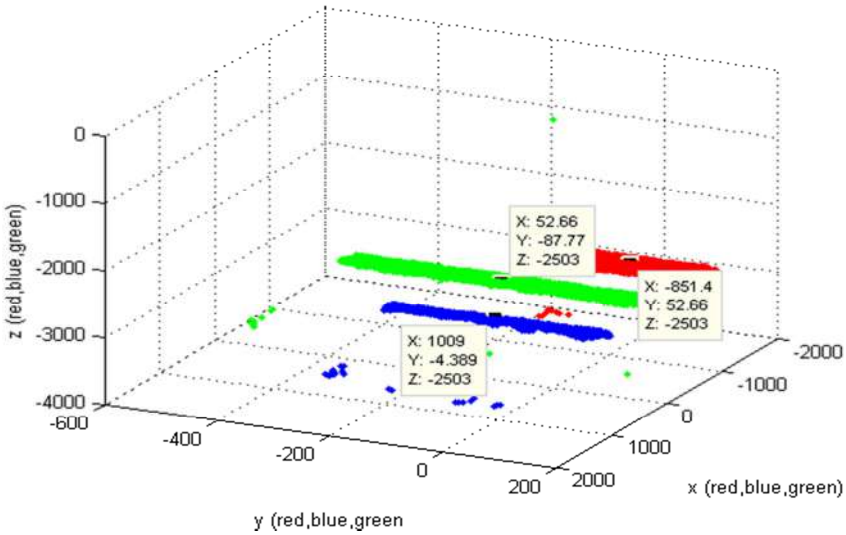


Figure IV.17 la segmentation 3D du premier essaie

Les statistiques métriques exploitent par la méthode de la segmentation 3D pour tous les objets entre eux et la kinect sont représenté dans le tableau suivant :

Tableau IV.2 Résultats des Distances Des objets Prisent Par Kinect (Exp1**)**

	rouge			Vert			Bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
kinect	851.4	4.38	2503	52.66	87.77	2503	1009	52.86	2503
bleu	1860.4	57.04	00	956.34	83.38	00			
vert	904.06	140.43	00				956.34	83.38	00
rouge				904.06	140.43	00	1860.4	57.04	00

Discutions des résultats

A partir de la segmentation 3D on peut mesurer les distance entre les objets ainsi la kinect on remarque que :

Selon l'axe (x) les résultats sont les suivant:

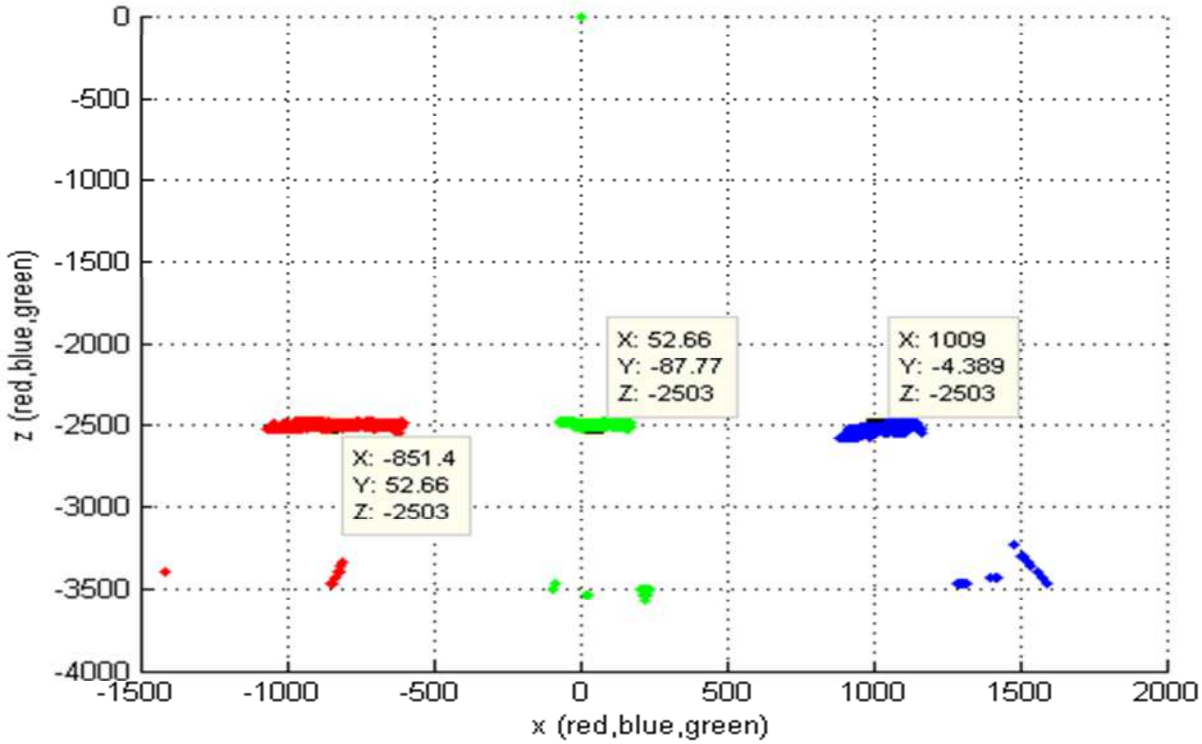


Figure IV.18 Représentation 3D des objets colorés selon l'axe (X)

- ✓ -Distance entre rouge et vert

$X=(X_{Vert}) – (X_{rouge})= (52.66) – (- 851.4) = 904.06$ mm

- ✓ -Distance entre rouge et bleu

$X=(X_{bleu}) – (X_{rouge})= (1009) – (- 851.4) = 1860.4$ mm

- ✓ -Distance entre vert et bleu

$X=(X_{bleu}) – (X_{vert})= (1009) – (52.66) = 956.34$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation 3D,les résultats sont représentés dans le tableau suivant

Tableau IV.3 : résultats d'erreur de la segmentation 3D suivants X(Exp1)

	rouge	Vert	Bleu
Erreur	X (mm)	X (mm)	X (mm)
kinect	48.6	42.66	9
bleu	119.6	23.66	
vert	95.94		23.66
rouge		95.94	119.6

On remarque qu'on a une erreur de segmentation $\Delta_x \leq 120$ mm

Selon l'axe (Y) les résultats sont les suivant:

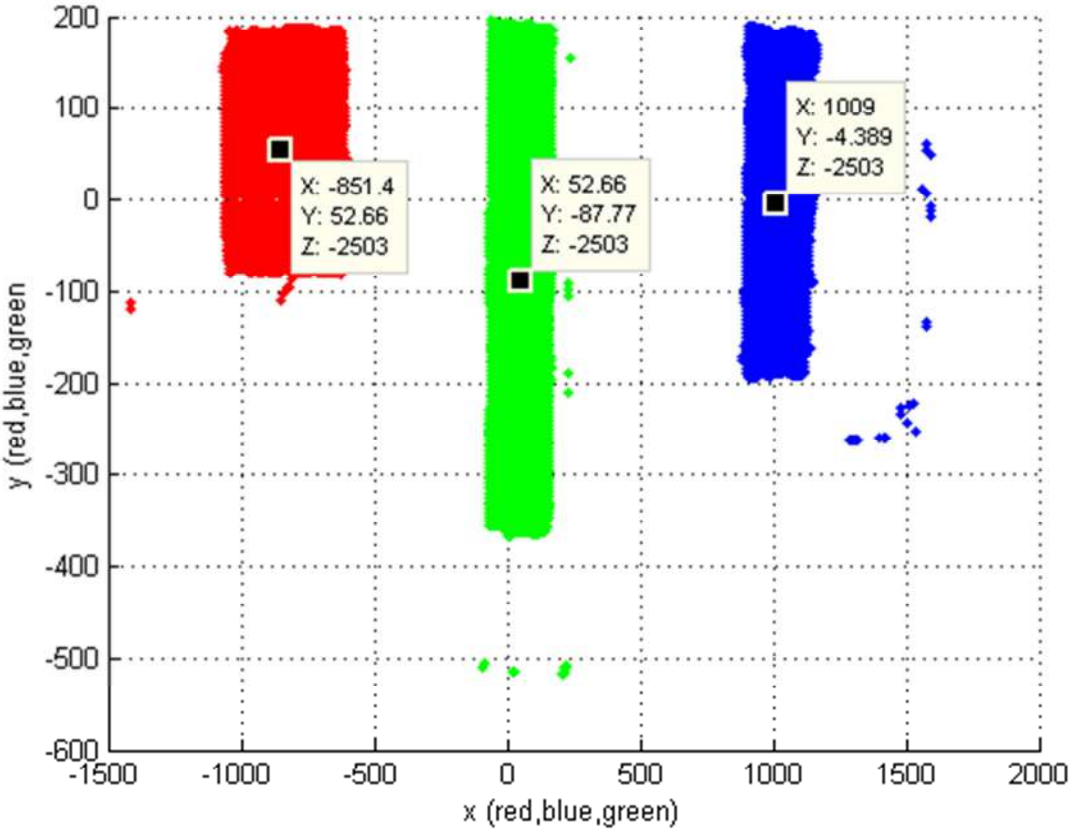


Figure IV.19Représentation 3D des objets colorés selon l'axe (Y)

- ✓ Distance entre rouge et vert

$Y= (Y_{rouge}) – (Y_{vert})= (52.66) – (- 87.77) = 140.43$ mm

- ✓ Distance entre rouge et bleu

$Y= (Y_{rouge}) – (Y_{bleu})= (52.66) – (- 4.389) = 57.04$ mm

- ✓ Distance entre vert et bleu

$Y= (Y_{bleu}) – (Y_{vert})= (1009) – (52.66) = 83.38$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant

Tableau IV.4 : résultats d'erreur de la segmentation 3D suivants Y (Exp1)

	rouge	Vert	Bleu
	Y (mm)	Y (mm)	Y (mm)
kinect	4.38	87.77	52.86
bleu	57.04	83.38	
vert	140.43		83.38
rouge		140.43	57.04

On remarque qu'on a une erreur de segmentation $\Delta_y \leq 140$ mm

Selon l'axe (Z) les resulat sont les suivant:

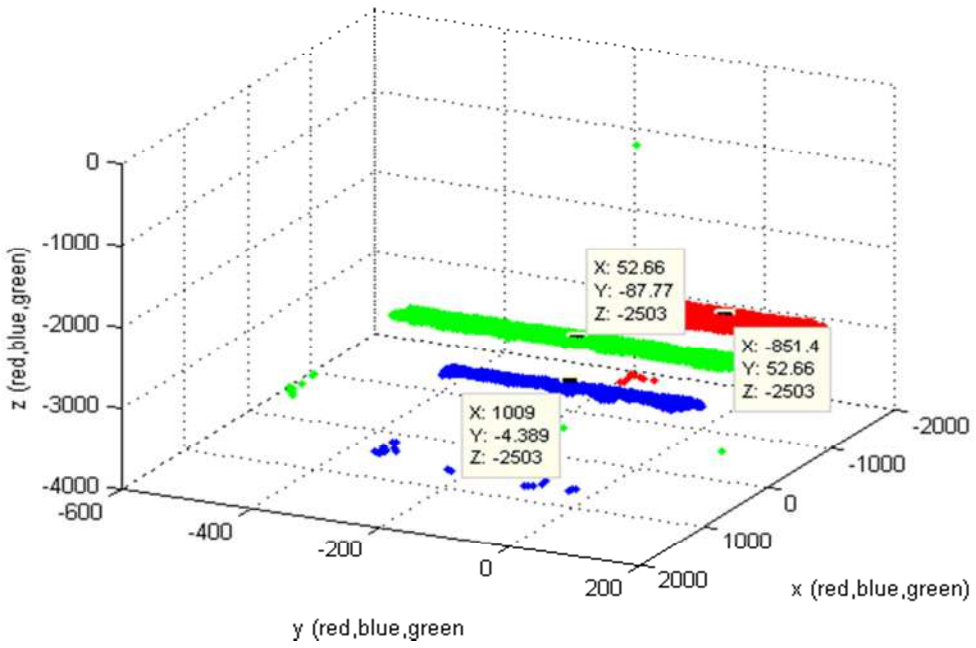


Figure IV.20Représentation 3D des objets colorés selon l'axe (Z)

- ✓ Distance entre rouge et vert

$Z=(Z_{rouge}) – (Z_{vert})= (2503) – (- 2503) = 0$ mm

- ✓ Distance entre rouge et bleu

$Z=(Z_{rouge}) – (Z_{bleu})= (2503) – (- 2503) = 0$ mm

- ✓ Distance entre bleu et vert

$Z=(Z_{bleu}) – (Z_{vert})= (2503) – (2503 = 0$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant :

Tableau IV.5 : résultats d'erreur de la segmentation 3D suivants Z(Exp1)

	rouge	Vert	Bleu
	Z (mm)	Z (mm)	Z (mm)
kinect	0.03	0.03	0.03
bleu	00	00	
vert	00		00
rouge		00	00

On remarque qu'on a une erreur de segmentation $\Delta_z \leq 0.05$ mm

Expérience 2 la segmentation suivent l'axe X.Y.Z: dans ce cas on a positionné les objets dans l'espace de telle façon qu'ils soient éloignés de différant distance par rapport à l'axe Y, X, Z de la Kinect.



Figure IV.21Image Original 2

La statistique métrique réelle de positionnement des objets (rouge, vert, bleu) entre eux, ainsi la kinect sont représenté par le tableau suivant:

Tableau IV.6 Distances des objets prisent manuellement (Exp2)

	rouge			vert			Bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
kinect	1000	0	3200	0	0	2500	500	0	1300
bleu	1450	50	2000	600	0	1150			
vert	850	50	780				600	0	1150
rouge				850	50	650	1450	50	2000

Notre algorithme de la segmentation 3D a fourni l'image suivante :

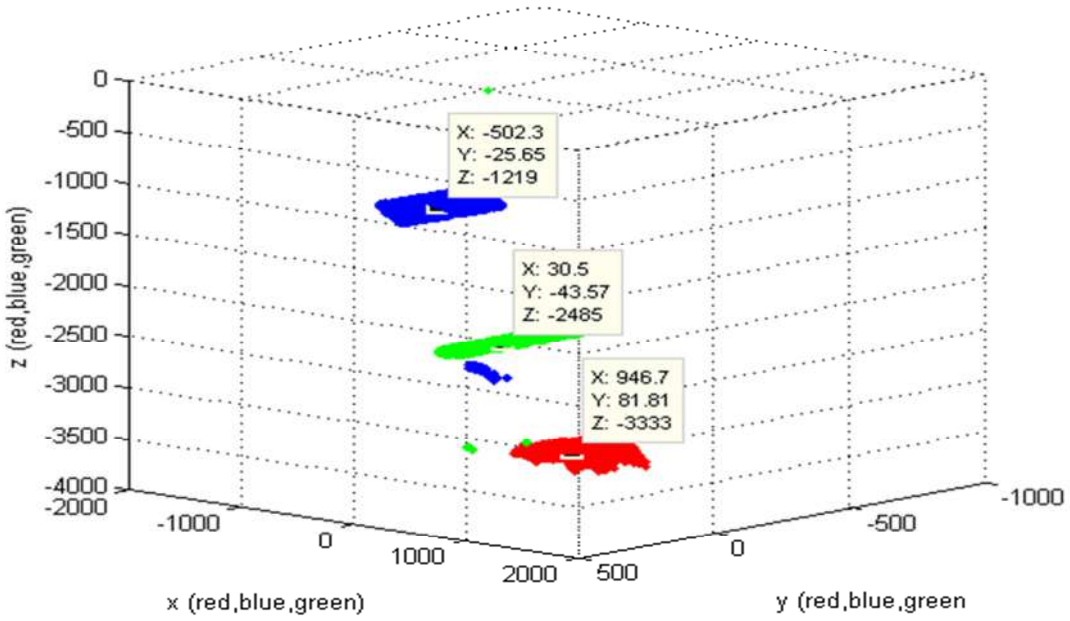


Figure IV.22 Représentation des objets segmenté en 3D

A partir de la segmentation 3D on peut extraire les statistiques de positionnement des objets dans l'espace, les résultats sont les suivant:

Tableau IV.7 Résultats des Distances Des objets Prisent Par Kinect (Exp2)

	rouge			vert			Bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
kinect	946.7	81.81	3333	30.5	43.57	2485	502.03	25.65	1219
bleu	1449	107.46	2114	532.8	17.92	1266			
vert	916.5	125.38	848				532.8	17.92	1266
rouge				916.5	125.38	848	1449	107.46	2114

Discutions des résultats:

A partir de la segmentation 3D on peut mesurer les distance entre les objet ainsi la kinect on remarque que

Selon l'axe (X) les résultat sont les suivant:

- ✓ Distance entre bleu et vert

$X=(X_{Vert}) – (X_{bleu})= (30.5) – (- 502.3) = 532.8$ mm

- ✓ Distance entre bleu et rouge

$X=(X_{rouge}) – (X_{bleu})= (946.7) – (- 502.3) = 1449$ mm

- ✓ Distance entre vert et rouge

$X=(X_{rouge}) – (X_{vert})= (946.7) – (30.5) = 916.5$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant :

Tableau IV.8 résultats d'erreur de la segmentation 3D suivants X(Exp2)

	rouge	Vert	Bleu
	X (mm)	X (mm)	X (mm)
kinect	53.3	30.5	2.3
bleu	1	67.2	
vert	66.5		67.2
rouge		66.5	1

On remarque qu'on a une erreur de segmentation $\Delta_X \leq 70$ mm

Selon l'axe (Y) les résultat sont les suivant:

- ✓ Distance entre bleu et vert

$Y= (Y_{bleu}) – (Y_{vert})= (-25.65) – (- 43.57) = 17.92$ mm

- ✓ Distance entre bleu et rouge

$Y= (Y_{bleu}) – (Y_{rouge})= (-25.65) – (- 4.389) = 107.46$ mm

- ✓ Distance entre vert et rouge

$Y= (Y_{vert}) – (Y_{rouge})= (1009) – (52.66) = 125.38$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant

Tableau IV.9 résultats d'erreur de la segmentation 3D suivants Y(Exp2)

	rouge	Vert	Bleu
Erreur	Y (mm)	Y (mm)	Y (mm)
kinect	81.81	43.57	25.65
bleu	57.4	17.92	
vert	75.3		17.92
rouge		75.3	57.4

On remarque qu'on a une erreur de segmentation $\Delta_Y \leq 76$ mm

Selon l'axe (Z) les résultat sont les suivant:

- ✓ Distance entre bleu et rouge

$Z= (Z_{bleu}) – (Z_{rouge})= (-1219) – (- 3333) = 2114mm$

- ✓ Distance entre bleu et vert

$Z= (Z_{bleu}) – (Z_{vert})= (1219) – (- 2485) = 1266$ mm

- ✓ Distance entre vert et rouge

$Z= (Z_{vert}) – (Z_{rouge})= (2485) – (-3333)= 848$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant:

Tableau IV.10 résultats d'erreur de la segmentation 3D suivants Z(Exp2)

	rouge	Vert	Bleu
	Z (mm)	Z (mm)	Z (mm)
kinect	100	15	81
bleu	114	116	
vert	68		116
rouge		68	114

On remarque qu'on a une erreur de segmentation $\Delta_z \leq 120$ mm

Expérience 3 la segmentation suivent l'axe Y:dans le troisième cas on a positionné les objets dans l'espace de telle façon qu'ils soient éloignés diffèrent distance par rapport à l'axe Y de la Kinect.

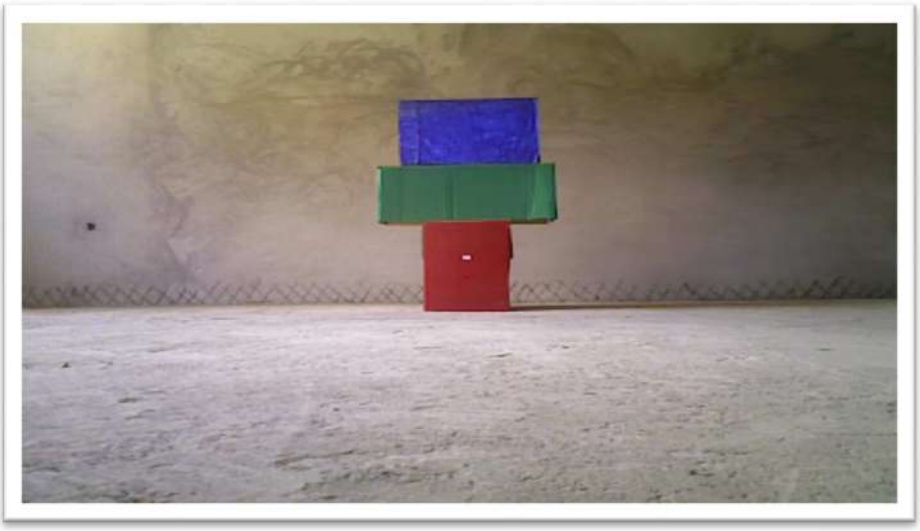


Figure IV.23Image Original 3

La statistique métrique réelle de positionnement des objets (rouge, vert, bleu) entre eux, ainsi la kinect sont représenté par le tableau suivant:

Tableau IV.11 Distances des objets prisent manuellement (Exp3)

	rouge			vert			Bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
kinect	0	600	2500	0	300	2500	0	0	2500
bleu	0	600	0	0	300	0			
Vert	0	300	0				0	300	0
rouge				0	300	0	0	600	0

Notre algorithme de la segmentation 3D a fourni l'image suivante :

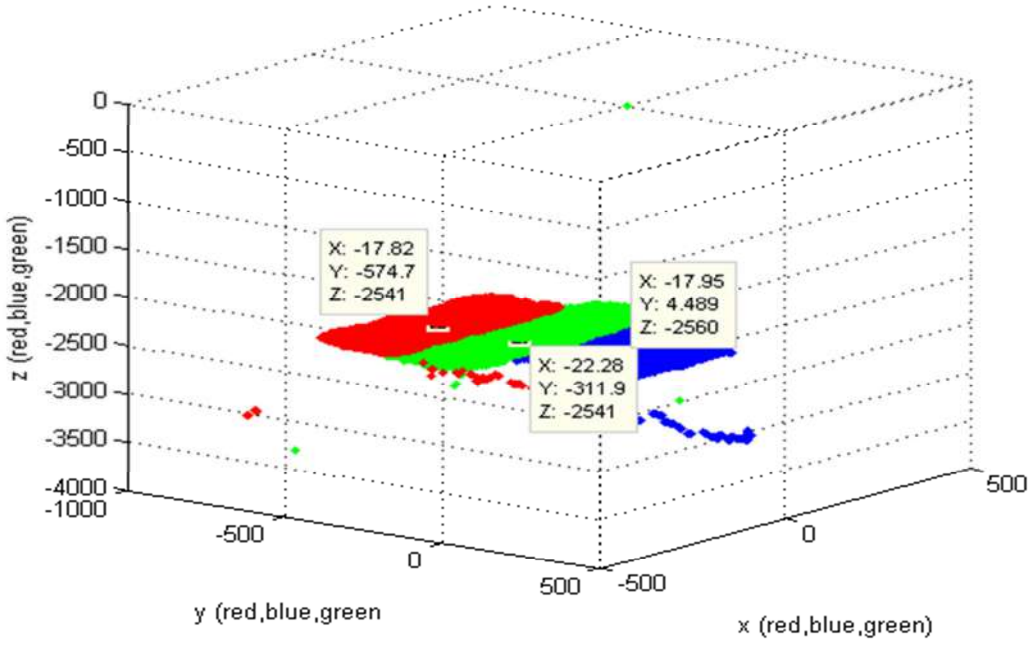


Figure IV.24 Représentation de la segmentation 3D

A partir de la segmentation 3D on peut extraire les statistique de positionnement des objets dans l'espace le résultat sont les suivant:

Tableau IV.12 Résultats des Distances Des objets Prisent Par Kinect (Exp3)

	rouge			vert			bleu		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
kinect	17.82	574.4	2541	22.28	311.9	2541	17.95	4.489	2541
bleu	0.13	579.18	0	4.33	316.38	19			
Vert	4.46	579.18	19				4.33	316.38	19
rouge				4.46	579.18	19	0.13	579.18	0

Discutions des résultats:

A partir de la segmentation 3D on peut mesurer les distance entre les objets ainsi la kinect on remarque que:

Selon l'axe (X) les résultat sont les suivant:

Distance entre rouge et bleu

$X=(X_{rouge}) - (X_{bleu}) = (-17.82) - (-17.95) = 0.13$ mm

Distance entre rouge et vert

$X=(X_{rouge}) - (X_{vert}) = (17.82) - (-22.28) = 4.46$ mm

Distance entre bleu et vert

$X=(X_{bleu}) - (X_{vert}) = (17.95) - (30.5) = 4.33$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant :

Tableau IV.13 résultats d'erreur de la segmentation 3D suivants X(Exp3)

	rouge	Vert	Bleu
	X (mm)	X (mm)	X (mm)
kinect	17.82	22.2	17.9
bleu	0.13	4.33	
vert	4.46		4.33
rouge		4.46	0.13

On remarque qu'on a une erreur de segmentation $\Delta_X \leq 25$ mm

Selon l'axe (Y) les résultat sont les suivant:

Distance entre bleu et vert

$Y= (Y_{bleu}) - (Y_{vert}) = (4.489) - (-311.9) = 316.389$ mm

Distance entre bleu et rouge

$Y= (Y_{bleu}) - (Y_{rouge}) = (4.489) - (-574.7) = 579.189$ mm

Distance entre vert et rouge

$Y= (Y_{vert}) - (Y_{rouge}) = (-311.9) - (574.7) = 262.8$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant :

Tableau IV.14 résultats d'erreur de la segmentation 3D suivants Y(Exp3)

	rouge	Vert	Bleu
	Y (mm)	Y (mm)	Y (mm)
kinect	24.3	11.9	4.489
bleu	20.9	16.3	
vert	37.2		16.3
rouge		37.2	20.9

On remarque qu'on a une erreur de segmentation $\Delta_y \leq 25$ mm

Selon l'axe (Z) les résultat sont les suivant:

- ✓ Distance entre rouge et vert

$Z = (Z_{bleu}) - (Z_{rouge}) = (-2541) - (-2541) = 0$mm

- ✓ Distance entre vert et bleu

$Z = (Z_{bleu}) - (Z_{vert}) = (2541) - (-2560) = 19$ mm

- ✓ Distance entre rouge et bleu

$Z = (Z_{vert}) - (Z_{rouge}) = (2541) - (-2560) = 19$ mm

A partir de ces données (mesure et exploité) on peut calculer l'erreur de la segmentation, les résultats sont représentés dans le tableau suivant :

Tableau IV.15 résultats d'erreur de la segmentation 3D suivants Z(Exp3)

	rouge	Vert	Bleu
	Z (mm)	Z (mm)	Z (mm)
kinect	0.41	0.41	0.60
bleu	19	19	
vert	0		19
rouge		0	19

On remarque qu'on a une erreur de segmentation $\Delta_z \leq 20$ mm

On générale notre algorithme de la segmentation 3D fourni de résultat avec une erreur inferieur ou égale a sa nous a permis de valider notre algorithme de segmentation 3D pour un enivrement publique

Expérience 4: dans ce cas on a pris un exemple souple



Figure IV.25 Image quatrième essaie

Le tableau suivant represent les valeur réel :

Tableau IV.16 Distances des objets prisent manuellement(Exp4)

	rouge			bleu		
	x (mm)	y (mm)	z (mm)	x (mm)	y (mm)	z (mm)
kinect	1200	100	2800	0	200	2500
bleu	1300	50	500			
rouge				1300	50	500



a b

Figure IV.26 Segmentation 2D :a)segmentation des objets rouges, b) segmentation des objets bleu

Apres la Reconstruction 3D de tous les l'objet on obtient le résultat:

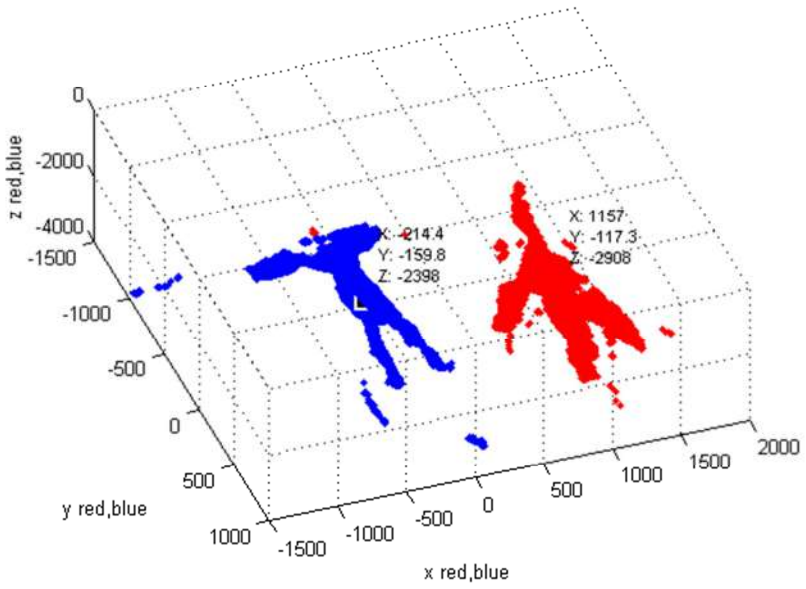


Figure IV.27 la segmentation 3D de la quatrième essaie

Les statistiques métrique exploite par la méthode de la segmentation 3D pour tous les objet entre eux et la kinect sont représenté dans le tableau suivant :

Tableau IV.17 Résultats des Distances Des objets Prisent Par Kinect (Exp4)

	rouge			bleu		
	x (mm)	y (mm)	z (mm)	x (mm)	y (mm)	z (mm)
kinect	1157	117.3	2908	214.4	159.8	2398
bleu	1371.4	42.5	510			
rouge				1371.4	42.5	510

Discutions des résultats

Selon l'axe (x) les résultat sont les suivant:

A partir de la segmentation 3D on peut mesurer les distance entre l'objet ainsi la kinect on remarque que.

- ✓ Distance entre rouge et bleu

$X=(X_{rouge)} – (X_{bleu})= (1157) – (-214.4) = 1371.4$ mm

Selon l'axe (Y) les résultat sont les suivant:

- ✓ Distance entre bleu et vert

$Y= (Y_{bleu)} – (Y_{rouge})= (-159.8) – (- 117.3) = 42.5$ mm

Selon l'axe (Z) les résultat sont les suivant:

..

- ✓ Distance entre rouge et bleu

$Z= (Z_{bleu}) – (Z_{rouge})= (-2908) – (- 2398) = 510$ mm

Dans les expériences qu'on a déjà fait précédemment, on a montré l'efficacité de notre algorithme quel que soit le positionnement de l'objet dans l'espace

Remarque :

Les interférences des couleurs sur les représentations de la segmentation 3D sont du à la réflexion des couleurs des objets avec le plan (le milieu).

IV.6 Conclusion :

Au cours de cette réalisation et étude on a montré l'utilité de la segmentation 3D, car à l'aide de la Kinect on a trouvé une liberté totale en matière expérimentale. En effet, on a réalisé plusieurs expériences qui ont prouvé l'efficacité de l'algorithme conçu pour le positionnement des objets coloré dans un milieu.

Cet algorithme peut être amélioré de tell façon qu'il soit exploitable pour d'autres couleur non seulement les trois couleurs principale. Et aussi selon les besoin ou l'objectif recherché.

Conclusion générale

Conclusion générale

Dans ce mémoire, nous avons fait un tour exhaustif des méthodes de Segmentation d'image RGBD où de nombreux progrès restent à faire car il n'existe pas encore d'algorithme assez robuste permettant de détecter les contours de n'importe quel type d'image

Une image contient une grande quantité d'information, dont la diversité se perçoit au nombre de façons dont on dispose pour la décrire : on pourrait en donner un aperçu à l'aide de l'orientation de ses contours principaux, ou bien en faisant la liste des objets qu'elle contient, classés selon leur netteté, ou encore en la découpant en régions, dont on décrit la couleur ou la texture...

La segmentation d'images constitue une phase importante du processus d'analyse d'image. L'évaluation de la segmentation est indispensable pour détecter les objets dans n'importe endroit, d'où nous souhaitions également caractériser les objets : dans ce but, nous avons utilisé la méthode de la segmentation par histogramme puis on utilise une autre méthode pour la reconstruction tridimensionnel

Donc dans notre travail on arrive à segmenter l'image RGBD telle façon on extraire chaque objet individuellement et ça nous permettre d'utilise notre algorithme dans des plusieurs domaine exemple : le robot, les nouvelle voiture, le domaine militaire …….

A la fin nous espérons que notre modeste travail a fourni peu de ca dans ce vaste domaine.

Bibliographie

Chapitre 1

[01] Jean-Philippe. Pernin –DIP ;'' l'image fixe - Les formats MATRICIELS (ou bitmap)'' ; Université Stendhal ; 2003

[02] G.M. COCHARD, S.CHOPLIN ;'' infographie et image informatique''.

[03] R.C. GONZAILES &P.WINTZ; ''Digital Image Processing''; AddisionWessley ; 1977.

[04] WILLIAM PUECH ; '' Vision par ordinateur'' ; Université Montpellier II- Nîmes.

[6] A.MANZANERA Traitement et reconnaissance d'images ,*Cours TERI – Master IAD UPMC Paris 6.* http://www.ensta-paristech.fr/~manzaner/Cours/IAD/AM_Filtrage.pdf

[07] M.KUNT ;''Traitement numérique des images'' ; Vol.2, 1993.

[08] K. TABARI & S. TAGMH ; ''Compression d'image animées à très débit par la géométrie des Fractales'' ; PEE USTHB ;1996

Chapitre2

[09] L.BUSIN 2006 : "sélection automatique d'espace couleur pour la segmentation d'images : application à la qualité de motifs de verres sérigraphies", thèse de doctorat, université de LILLE 1 France.

[10] A.porebski, N.Vandenbroucke, L.Macaire : "Attributs de texture couleur", article, (EIPC, LAGIS, UMR), France

[11] L. PAULHAC 2009 : "outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques", thèse de doctorat, université François Rabelais de tous France.

[12] S.OUATARA 2009 : "stratégies de segmentation d'images multi-composantes par analyse d'histogrammes multidimensionnels", thèse de doctorat, école doctorale d'Angers STIM, France

[13] R. C. Gonzalez et al 2002: "Digital Image Processing Using Matlab", livre, USA.

[14] M. Fontaine 2001 : "segmentation non supervisée d'images couleur par analyse de la connexité des pixels", thèse de doctorat, université de Lille 1, France

[15] L.Macaire 2004: "exploitation de la couleur pour la segmentation et l'analyse d'images", mémoire présenté en vue de l'obtention de l'habilitation à UFR d'I.E.E.A, France.

[16] L.Busin 2003 : “segmentation d’images par analyse récursive d’histogrammes couleur monodimensionnels“, rapport de stage, université des sciences et technologies de Lille et école centrale de Lille, France

Chapitre 3

[17] Radu Horaud et OulivierMonga , vision par ordinateur 1993(first edition) [ISI 98] ISIMA 1988,1999. Introduction a` C++ builder

[18] .Barsi, L. Gosta, D.GeigeretD.Jacobs. »Determining the similarity of deformable shapes» 1998.

[19] R. Zhang, P.-S. Tsai, J. Cryer and M. Shah. Shape from shading : A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, August 1999. Webliographie,

[20] R. Kozera » An overview of the shape from shading problem» Machine Graphics and Vision, 1998.

[21] Pierre-Marc Jodoin »Vision par ordinateur» Hiver 2009

[22] . Blum. A Transformation for Extracting New Descriptors of Shape, In Models for the Perception of Speech and Visual Form, (W.Wathen-Dunn, ed.), Cambridge Press, 1967

[23] .Bubel, R. Bergevin, Classifying junctions by vector quantization, in : Vision Interface'98, 1998.

[24] F.Hetitger.Feature detection using suppression and enchancement. Technical Report 163.Communication Technology Laboratory.Swiss Federal Institute of Technology ITH 1995.

[26] J.Coughlan, A.Yuille,C.English et D.Snow » Efficient deformable template detection and localization without user initialization» 2000.

http://www.memoireonline.com/09/10/3915/m_Reconnaissance-des-objets-polyedriques3.html

Chapitre 4

(http://www.gameblog.fr/article-lecteur_612_comment-fonctionne-la-technologie-kinect)
(Docteur Sub)

http://www.asus.com/Multimedia/Xtion_PRO_LIVE/