

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE RECHERCHE SCIENTIFIQU



جامعة محمد بوضياف - المسيلة
UNIVERSITE MOHAMED BOUDIAF - M'SILA



كلية التكنولوجيا
FACULTE DE TECHNOLOGIE
قسم الإلكترونيك
DEPARTEMENT D'ELECTRONIQUE

MEMOIRE DE MASTER

DOMAINE : SCIENCES ET TECHNOLOGIE
FILIERE : Génie Électrique
OPTION : Contrôle industriel

THEME

*Extraction de texte dans les séquences
vidéo*

Proposé et dirigé par :

Mr. LALAOUI LAHOUAOUI

Réalisé par :

HEBIL Salah

N° D'ordre: **2016/CI_N°Sujet /CI15**

PROMOTION: JUIN 2016

Remerciement

Je tiens tout d'abord à remercier ALLAH le tout puissant et miséricordieux, qui m'a donné la force et la patience d'accomplir ce mémoire.

En second lieu, Je tiens à remercier particulièrement mes parents ; mon succès demeure de loin le fruit de leurs longues années de sacrifices et de patience.

Mon vifs remerciements, accompagnés de toute mon gratitude, vont tout d'abord à mon encadreur Mr. LALAOUI LAHOUAOUI, pour m'avoir proposé ce sujet et dirigé mon travail, pour toute leur aide, Je suis ravi d'avoir travaillé en ses compagnies car outre leur appui scientifique, ils sont toujours été là pour me soutenir et me conseiller au cours de l'élaboration de ce mémoire. Les conseils qu'ils m'ont prodigué, la patience, la confiance qu'ils m'ont témoignés ont été déterminants dans la réalisation de ce travail de recherche.

J'exprime ma reconnaissance aux membres de jury pour bien vouloir accepter de présider et examiner ce travail.

Nos remerciements s'étendent également à tous, nos enseignants durant les années des études.

Enfin, je remercie tous ceux qui ont contribué de près ou de loin, pour Leurs soutiens moraux ou matériels, dans la réalisation de ce projet.

Dédicaces

Je dédie ce modeste et mémorable travail, plus particulièrement à mes aimables et respectueux parents qui m'ont accompagné dans toute ma vie, s'inquiétant énormément pour m'offrir une vie meilleure. Je tiens la parole pour leur dire : « Voilà les fruits de vos sacrifices ! ».

*À mes sœurs ; À toute La famille HEBIL et KOUIDRI;
À mes amis, et toute la promotion d'électronique et spécialement le groupe d'option contrôle Industriel ; Ainsi qu'à toutes les personnes qui m'a aidé à la réalisation de ce travail.*

Hebil Salah

Table des matières

Introduction générale	1
Chapitre I : Généralité sur traitement d'image	
I.1. Introduction	3
I.2. Définition de l'image	3
I.3. Image numérique.....	4
I.4. Les formats d'images	5
I.4.1. Windows BitMaP - BMP	5
I.4.2. Tagged Image File Format - TIFF.....	6
I.4.3. Joint Photographic Expert Group - JPEG	6
I.4.4. Graphics Interchange Format - GIF	6
I.4.5 . Portable Network Graphic - PNG	6
I.5. Caractéristiques de l'image	7
I.5.1. Pixel.....	7
I.5.2. Dimension, Résolution , La taille d'une image.....	8
I.5.2.1. Dimension.....	8
I.5.2.2. Résolution	8
I.5.2.3. La taille d'une image.....	9
I.5.3. Voisinage.....	9
I.5.4. Niveau de gris.....	10
I.5.5 Image couleur (RGB)	11
I.5.6. Contraste.....	13
I.5.7. Luminance	14
I.5.8. Bruit.....	15
I.5.9. Contour.....	15
I.5.10. Histogramme de l'image	16
I.6. Types d'image	17
I.6.1. Images matricielles (ou image bitmap)	17
I.6.1.1 . Images 2D.....	17
I.6.1.2. Image 2D +t (vidéo), image 3D, image multi-résolution	18
I.6.1.3. Images stéréoscopique	18

I.6.2. Images vectorielles	18
I.7. Traitement d'image.....	20
I.7.1. Définition de Traitement	20
I.7 les techniques fondamentales en traitement d'images.....	21
I.7.1 Acquisition de l'image	21
I.7.2 Le rehaussement de l'image	21
I.7.3 Les ondelettes	21
I.7.4 La morphologie	21
I.7.5 La segmentation	22
I.7.6 La représentation et description	22
I.7.7 La reconnaissance :	22
I.7.8 La restauration d'image.....	22
I.7.2. Historique	23
I.8 Conclusion	24

Chapitre II : Segmentation d'image

II.1 Introduction.....	25
II.2. Définition de la segmentation d'image.....	29
II.3. Les différentes approches de segmentation	30
II.3.1. Segmentation basée sur les contours.....	31
II.3.1.1 Définition (un contour) :.....	31
II.3.1.2 Méthode des contours actifs (snacke)	31
II.3.1.3 Méthode d'ensembles de niveau zéro (LEVEL SETS).....	32
II.3.2. Approches de segmentation par région	33
II.3.2.1. Définition (Segmentation par région).....	34
II.3.2.1.1 Méthodes par croissance de régions	34
II.3.3. Segmentation basée sur les pixels.....	38
II.2.3.1 Seuillage de l'image	38
II.2.3.2 Segmentation de vidéo	41
II.4 Conclusion	46

Chapitre III : Processus de détection de texte sur les séquences vidéo et les critères dévaluations

III.1	Introduction.....	47
III.2	Formulation du problème	58
III.3	Etat de l'art.....	59
III.4	Un système d'extraction	51
III.5	Les Critères d'évaluations	52
III.5.1	Avec segmentation de référence.....	53
III.5.1.1	Mesure de Vinet.....	54
III.5.1.2	Mesure de cohérence entre segmentations de Martin.	54
III.5.1.3	Position des pixels mal segmentés : mesure de Yasnoff et al.	55
III.5.2	Sans segmentation de référence.....	55
III.5.2.1	Contraste inter-région de Levine et Nazif.....	56
III.5.2.2	Contraste de Zeboudj	56
III.5.2.3	Critère de Rosenborger	57
III.6	Les operateurs de segmentation	57
III.6.1	Technique du Gradient	58
III.6.2	Opérateur de Robert.....	59
III.6.3	Opérateur de Sobel.	60
III.6.4	Operateur de Prewitt.....	62
III.6.5	Le Laplacien	62
III.6.6	Opérateur de Kirsh.....	63
III.6.7	Critère de Canny.....	63
III.6.8	Les détecteurs de passage par zéro du Laplacien	64
III.7	Conclusion.....	65

Chapitre IV : Expérimentation & validation

IV.1	Introduction.....	65
IV.2	Résultat des critères.....	70

IV.2.1	Nous avons fixe Sobel	71
IV.2.2	Nous avons fixe l'image Prewitt.....	72
IV.2.3	Nous avons fixe l'image Canny	73
IV.2.4	Nous avons fixes l'image Robert	74
IV.2.5	Nous avons fixe l'image de Niveau de gray.....	74
IV.3	Discussion des résultats.....	75
IV.5	Conclusion	76
	Conclusion générale	77
	Bibliographe	
	Résumi	

Liste des figures

Figure I.1	: Exemple d'une image numérique.	5
Figure I.2.	: Représentation d'un pixel.	8
Figure I. 3.	: Exemple de voisinage, (a) est voisinage à 4 et (b) voisinage à 8.	10
Figure I.4	: Image en niveau de gray.	10
Figure I.5	: Représentée l'image en couleurs 8 bits et la palette 8 bits.....	10
Figure I.6 – Palette 24 bits	Exemple de couleur en 24 bits	11
Figure I.7 – Image en couleurs 24 bits	12
Figure I.8	: Pour une valeur de + 96, tous les points de l'espace V' seront blancs.	13
Figure I.9	: Changer la luminosité dans l'image (a) et (b).	14
Figure I.10	: Exemple sur l'image sans bruit et bruitée.....	14
Figure I.11	: Détection de contour sur l'image Chever	15
Figure I.12	: Image avec histogramme.....	16
Figure I.13	: Exemple d'une image matriciel.	16
Figure I.14	: exemple d'une image vectorielle par logiciel Auto CAD.	17
Figure I.15	: Traitements fondamentaux en traitement d'images.	19

Figure II.1 – Segmentation : image originale.	21
Figure II.2 -Extraction des contours d’une image à partir de ses gradients.	27
Figure II.3 : Segmentation en régions à des contours morphologiques par watershed.	27
Figure II.4 –Segmentation en régions de couleur homogène.	28
Figure II.5 : Segmentation par contour actifs : du contour initial (à gauche) au contour à convergence (à droite).	28
Figure II.6 : principe de contour actif.	29
Figure II.7 : principe de Level Sets.	32
Figure II.8 : Le processus d’un algorithme de croissance de région. (a) début de processus, le point noir présente le germe et les flèches la direction de croissance. (b) la croissance de pixels après quelques itérations. (c) le résultat de la segmentation.	33
Figure II.9 : L'application de l'algorithme croissance de région sur les images mammographiques.	34
Figure II.10 : Principe de la Ligne de Partage des Eaux.	35
Figure II.11 : L'application de l'algorithme K-means.	36
Figure. II.12 : Détections de contours sur l’image de Léna pour différentes valeurs de seuils (a) $s_h = 29$ et $s_b = 26$, (b) $s_h = 0$ et $s_b = 0$, (c) $s_h = 60$ et $s_b = 55$, (d) $s_h = 32$ et $s_b = 30$	37
Figure II.13 : (a) Image originale des poivrons et détection de contours sur cette image avec (b) le même seuil que pour Léna $s_h = 29$ et $s_b = 26$ et (c) avec un seuil plus faible $s_h = 12$ et $s_b = 10$	39
Figure II.14 (a) Image originale bruitée avec un bruit de Poisson. (b) $s_h = 10$ et $s_b = 8$, (c) $s_h = 20$ et $s_b = 18$, (d) $s_h = 25$ et $s_b = 23$, (e) $s_h = 108$ et $s_b = 105$	40
Figure II.15 : En premier Ligne: L'échantillon entré des cadres de l'image de marple13 classe par [15].	41
Figure II.16 : Tranches d'un volume de la vidéo. La ligne du sommet montre quelques cadres de l'échantillon d'une vidéo.	43
Figure II.17 : le filtre Deriche est appliqué sur un volume de la vidéo pour extraire le premier plan (Blanc les images élémentaires constituent le premier plan).	45
Figure III.1 : Images d'exemples.	46
Figure : III.2 : Texte sur un fond complexe (a) un essai de seuillage fixe (b) Rehaussement artificiel du contraste (c) un essai de seuillage fixe (d)	47

Figure III.3 : Le fonctionnement général du système	49
Figure III.4 : Image détecter par filtre de roberts	51
Figure III.5 Sobel Bord Découverte Filtres.....	59
Figure III.6 : Filtre Sobel Vertical.....	61
Figure III.7 : Filtre Sobel Horizontal.	61
Figure III.8: Prewitt edge detection filter.	62
Figure III.9 : Détection de contours par Laplace.	63
Figure III.10 : Détection de contour par Canny.	64
Figure III.11 : Détection de contour passage par zéro.	65
Figure IV.1 : (a) Image originale, (b) l'image détectée par Sobel.	66
Figure IV.2 : Histogramme de l'image de l'image (b).....	66
Figure IV.3 : (c) Image original, (d) Image détection de Prewitt.	67
Figure IV.4 : Histogramme de l'image (d).	67
Figure IV.5 : (e) Image original, (f) Image détection de Canny.....	68
Figure IV.6 : Histogramme de l'image (f).	68
Figure IV.7 : (g) Image original, (h) Image détection de Robert.	69
Figure IV.8 : Histogramme de l'image (h).	69
Figure IV.9 : (i) Image original, (j) Image en Niveau de gray.	70
Figure IV.10 : Histogramme de l'image de l'image (j).	70
Figure IV.11 : (a) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (b) est l'histogramme de tableau.	71
Figure IV.12 : (c) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (d) est l'histogramme de tableau.	72
Figure IV.13 : (e) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (f) est l'histogramme de tableau.....	73
Figure IV.14 : (g) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (h) est l'histogramme de tableau.	74
Figure IV.15 : (i) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (j) est l'histogramme de tableau.	75

Liste des Tableaux

Table I.1 : Tableau du codage d'une image en couleurs 24 bits.....	12
Tableau III.1 : Masques utilisés par Roberts Operateur.	59
Tableau III.2 : Masques utilisés par Opérateur de Sobel.....	60
Tableau III.3 : L'étiquetage d'images élémentaires de voisinage.	60
Tableau III.4 : Masques utilisés par Prewitt inclinaison Opérate.	62
Tableau IV.1 : représenté les résultats des opérateurs utilisés.	65

Introduction générale

Introduction générale

Le traitement d'images voit ses débuts dans les années 1920 dans la transmission de données par câble mais ne connaît de vrai essor que dans les années 1960 avec le développement des ordinateurs. Au départ, les techniques de traitement d'images sont essentiellement des méthodes de restauration et de compression d'images. Puis se développent avec les progrès de l'informatique, des techniques de détection de primitives (contours, point d'intérêt, lignes d'intérêt ...) et de nombreux autres traitements dans les domaines aussi variés que le médical, la télévision, l'imagerie satellitaire, le multimédia. C'est dans les années 2000 que l'image numérique et par conséquent le traitement d'images devient omniprésent. Que cela soit sur internet, au cinéma, à la télévision, sur les téléphones, dans le domaine médical, l'image est partout. Aujourd'hui il ne s'agit plus uniquement de traiter les images pour les améliorer mais aussi de les comprendre et de les interpréter. C'est dans ce contexte que la reconnaissance d'objets dans les images devient un sujet de recherche important. Et pour reconnaître des objets afin d'interpréter les images, il faut souvent au préalable les segmenter, c'est-à-dire séparer les objets d'intérêt du fond de l'image.

Comment déterminer les objets d'intérêt d'une image. C'est tout le problème de la segmentation d'images qui est un problème mal posé. La segmentation d'images numériques consiste à regrouper ensemble des pixels qui possèdent les mêmes caractéristiques. Il faut au préalable déterminer ce qui représente au mieux les objets. Cela peut être une couleur particulière, une forme précise. Mais est-ce suffisant. Pour l'humain, l'extraction des objets du fond de l'image repose sur un ensemble de critères et sur le fait qu'il reconnaît l'objet qui l'intéresse. Une part importante dans l'extraction des objets repose sur l'acquis de l'humain qui connaît les propriétés des objets présents dans la scène. Il sait (par expérience) par exemple qu'une balle est un objet de forme sphérique ou ovoïde et qu'un arbre possède en général un tronc et des feuilles (ou des épines) et qu'en hiver certains arbres n'ont plus de feuilles mais restent néanmoins des arbres. Comment reproduire cela avec un ordinateur qui ne dispose que de pixels.

Donne notre travail nous allons nous intéresser principalement aux images des vidéos contiennent un ou plusieurs objets (par ex : texte) sont séparables du reste de l'image.

Ce mémoire est divisé en quatre chapitres : tout d'abord nous présentons le cadre théorique de nos travaux, puis nous appliquons ce cadre général à la segmentation d'images et enfin à la segmentation de séquences vidéo.

Dans le premier chapitre nous avons présentés le cadre théorique de nos travaux ces des généralités sur le traitement d'images.

Le deuxième chapitre présente les différentes méthodes de segmentation basées sur les contours.

Dans le troisième chapitre nous avons présentés quelques critères de segmentation avec référence au sans références pour évaluer les résultats de segmentation.

Enfin le quatrième chapitre est consacré pour représenter les résultants de différentes méthodes avec les critères de validations.

Chapitre I

Généralité sur traitement d'image

I.1. Introduction

Avec la parole, l'image constitue l'un des moyens les plus importants qu'utilise l'homme pour communiquer avec autrui. C'est un moyen de communication universel dont la richesse du contenu permet aux êtres humains de tout âge et de toute culture de se comprendre. C'est aussi le moyen le plus efficace pour communiquer, chacun peut analyser l'image à sa manière, pour en dégager une impression et d'en extraire des informations précises.

De ce fait, le traitement d'images est l'ensemble des méthodes et techniques opérant sur celles-ci, dans le but de rendre cette opération possible, plus simple, plus efficace et plus agréable, d'améliorer l'aspect visuel de l'image et d'en extraire des informations jugées pertinentes.

Dans ce cadre s'inscrit ce projet qui porte sur le développement d'une application permettant d'intégrer sous une seule interface graphique des outils de traitement et de manipulation d'images. Cette application sera développée sous Matlab à l'aide du toolbox GUIDE. [1]

I. 2. Définition de l'image

L'image est une représentation d'une personne ou d'un objet par la peinture, la sculpture, le dessin, la photographie, le film, etc. C'est aussi un ensemble structuré d'informations qui, après affichage sur l'écran, ont une signification pour l'œil humain.

Représentation des images numériques :

Une image numérique est une matrice de pixels repérés par leur coordonnées (x, y).

S'il s'agit d'une image couleur, un pixel est codé par 3 composantes (r, g, b) (chacune comprise au sens large entre 0 et 255), représentant respectivement les "doses" de rouge, vert et bleu qui caractérisent la couleur du pixel. S'il s'agit d'une image en niveau de gris, il est

codé par 1 composante comprise au sens large entre 0 et 255, représentant la luminosité du pixel.

I.3. Image numérique

Contrairement aux images obtenues à l'aide d'un appareil photo, ou dessinées sur un papier, les images manipulées par un ordinateur sont numériques (représentées par une série de bit).

L'image numérique est l'image dont la surface est divisée en éléments de taille fixe appelés cellules ou pixels, ayant chacun comme caractéristique un niveau de gris ou de couleurs prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description interne de la scène à représenter.

La numérisation d'une image est la conversion de celle-ci de son état analogique en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques $f(x, y)$, comme la montre la figure 1.4 où :

x, y : coordonnées cartésiennes d'un point de l'image.

$F(x, y)$: niveau d'intensité.

Pour des raisons de commodité de représentation pour l'affichage et l'adressage, les données images sont généralement rangées sous formes de tableau I de n lignes et p colonnes.

Chaque élément $I(x, y)$ représente un pixel de l'image et sa valeur est associée à un niveau de gris codé sur m bits

(2^m niveaux de gris ; 0 = noir ; $2^m - 1$ = blanc).

La valeur 0 en chaque point exprimer la mesure d'une intensité lumineuse perçue par le Capteur.

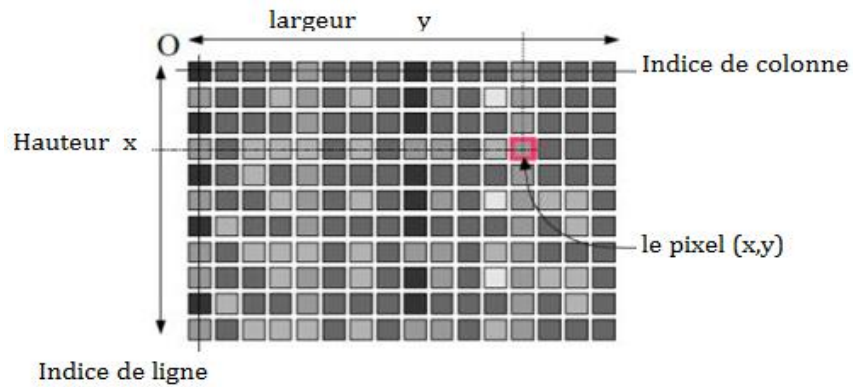


Figure I.1 : Exemple d'une image numérique. [1]

I.4. Les formats d'images

A l'instar des fichiers musicaux ou vidéo, les fichiers images disposent eux aussi de plusieurs formats. Les 5 formats que je vais détailler, qui sont également les plus utilisés, sont BMP, TIFF, JPEG ou JPG, GIF et PNG.

A l'instar des fichiers musicaux ou vidéo, les fichiers images disposent eux aussi de plusieurs formats.

Les 5 formats que je vais détailler, qui sont également les plus utilisés, sont :

- BMP : Windowes BitMap.
- TIFF: Tagged Image File Format.
- JPEG: Joint Photographic Expert Group.
- GIF : Graphics Interchange Formlat.
- PNG : Portabel Network Graphic.

I.4.1. Windows BitMaP - BMP

Le format BMP est un des premiers formats d'image utilisé sous Windows. Il est un des seuls formats à ne pas utiliser.

Cette technologie à pour principal avantage la qualité des images fournies : pas de compression = pas de perte de qualité.

Revers de la médaille : ne pas compresser le fichier donne des fichiers de grande taille, impossible à afficher sur internet pour un utilisateur ayant une connexion bas débit.

Pour exemple, une image 800x600 pixels pèsera 1.37Mo.

De plus, les fichiers n'étant pas compressés, deux images ayant la même définition auront toujours la même taille.

I.4.2. Tagged Image File Format - TIFF

Ce format est orienté vers les professionnels (imprimeurs, publicitaires...) car il a l'avantage d'être reconnu sur tous types de système d'exploitation : Windows, Mac, Linux, Unix ...etc.

Il permet d'obtenir une image de très bonne qualité, mais sa taille reste volumineuse, même si elle est inférieure à celle des fichiers BMP.

I.4.3. Joint Photographic Expert Group - JPEG

Ce format offre des taux de compression inégalés, même si la qualité de l'image s'en ressent au fur et à mesure que vous augmentez la compression.

Avec des taux de compression élevés donnant lieu à des fichiers images de petite taille, ce format est devenu le standard des formats d'image sur internet.

En effet, des fichiers de petites tailles seront chargés rapidement, même par une connexion basse débit.

I.4.4. Graphics Interchange Format - GIF

Ce format est l'autre standard d'internet. Les fichiers GIF sont de petites tailles, ce qui est dû au fait que ces images ne peuvent enregistrer que 256 couleurs : le plus gros avantage du format est lié à son plus gros inconvénient. Le format GIF permet également la création d'animations et de détourage.

I.4.5. Portable Network Graphic - PNG

C'est le format appelé à devenir le futur standard internet. Comme le GIF il permet le détourage des images, mais là où le format GIF enregistre 256 couleurs, le PNG en retient

16.7 Millions ce qui offre une image parfaite, avec un excellent rendu des nuances et des dégradés.

La taille des fichiers reste raisonnable, et, technologie dont ce format est le seul à disposer, il permet la compression sans perte de donnée ! C'est donc le format en devenir.

I.5. Caractéristiques de l'image

L'image est un ensemble structuré d'information caractérisé par les paramètres suivants :

- 1. Pixel
- 2. Démentions, Résolution Et La Taille D'image
- 3. Voisinage
- 4. Niveau de gris
- 5. Image couleur (RGB)
- 6. Contraste
- 7. Luminance
- 8. Bruit
- 9. Contour
- 10. Histogramme de l'image

I.5.1. Pixel

Le pixel est l'abréviation du mot « Picture élément », est une unité de surface permettant de définir la base d'une image numérique. Il matérialise un point donné (x, y) du plan de l'image. L'information représentée par le pixel est le niveau de gris (ou la couleur) prélevée à l'emplacement correspondant dans l'image réelle.

La différence entre image monochrome et image couleur réside dans la quantité d'informations contenue dans chaque pixel, par exemple dans une image couleur (RVB : Rouge, Vert, Bleu) la valeur d'un pixel est représentée sur trois octets pour chaque couleur.

[1]

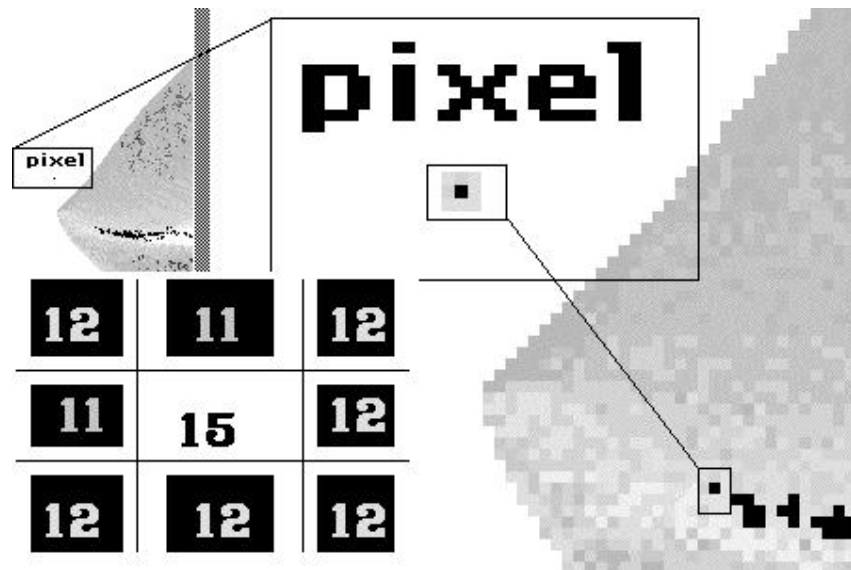


Figure I.2. : Représentation d'un pixel.

I.5.2. Dimension, résolution et la taille d'image

I.5.2.1. Dimension

C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image.

I.5.2.2. Résolution

La résolution est définie par un nombre de pixels par unité de longueur de l'image à numériser en dpi (dots per inch) ou ppp (points par pouce). On parle de définition pour un écran et de résolution pour une image. Plus le nombre de pixels est élevé par unité de longueur de l'image à numériser, plus la quantité d'informations qui décrit l'image est importante et plus la résolution est élevée (et plus le poids de l'image est élevé).

La résolution d'une image correspond au niveau de détail qui va être représenté sur cette image. Pour la numérisation il faut considérer les 2 équations suivantes :

$$(x * \text{résolution}) = x \text{ pouce}$$

$$(Y * \text{résolution}) = y \text{ pouce}$$

Telle que :

- **X** et **Y** représenté la taille (pouce ou cm, **1 pouce = 2,54** centimètres) de la structure à numériser.
- résolution représente la résolution de numérisation.
- **X** et **Y** représenté la taille (en pixels) de l'image.

I.5.2.3. 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.

Exemple :

Pour une image de **240*240** en Vraie couleur :

Nombre de pixel : $240 * 240 = 100800$,

Et la Taille de chaque pixel est : **24 bits / 8 = 3 octets**

Le poids de l'image est ainsi égal à :

$$100800 * 3 = 302.400 \text{ égal } 302.400/1024=395 \text{ ko}$$

I.5.3. Voisinage

Le plan de l'image est divisé en termes de formes rectangulaires ou hexagonales permettant ainsi l'exploitation de la notion de voisinage. [2]

Le voisinage d'un pixel est formé par l'ensemble des pixels qui se situent autour de ce même pixel. On définit aussi l'assiette comme étant l'ensemble de pixels définissant le voisinage pris en compte autour d'un pixel.

On distingue deux types de voisinage :

- Voisinage à 4 : On ne prend en considération que les pixels qui ont un coté commun avec le pixel considéré.

- Voisinage à 8 : On prend en compte tous les pixels qui ont au moins un point en liaison avec le pixel.

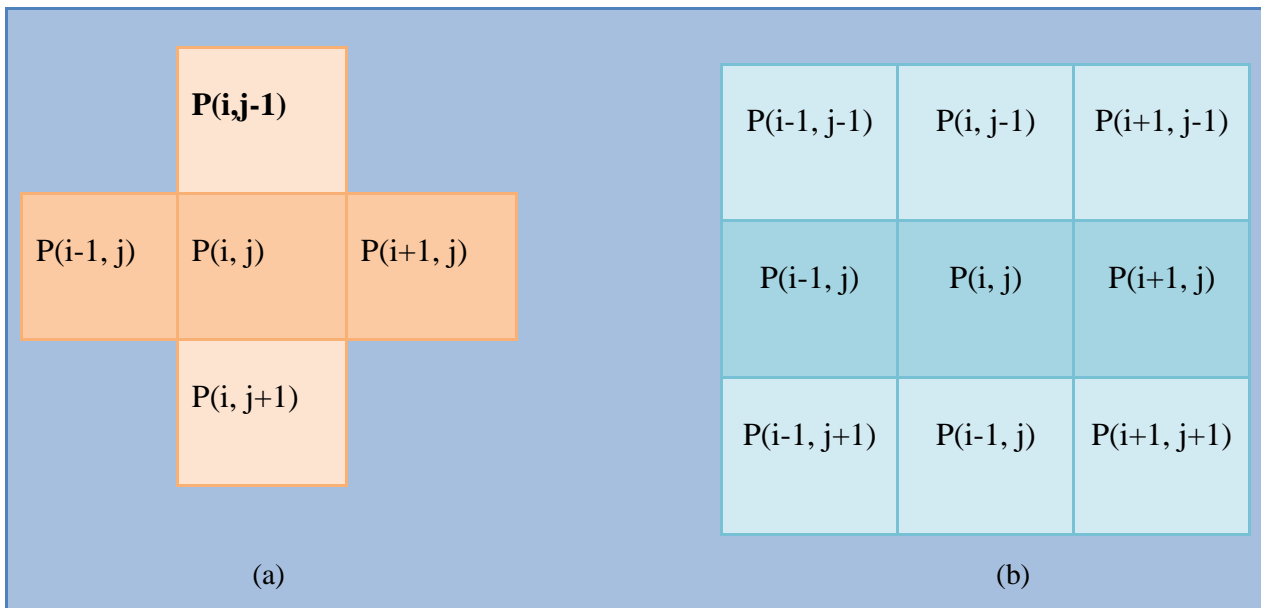


Figure I. 3. : Exemple de voisinage, (a) est voisinage à 4 et (b) voisinage à 8. [3]

I.5.4. Niveau de gris

C'est la valeur d'intensité lumineuse d'un pixel. Cette valeur peut aller du noir (0) jusqu'au blanc (255) en passant par les nuances qui sont contenues dans l'intervalle [0, 255]. Elle correspond en fait à la quantité de la lumière réfléchie.

Pour 8 bits, on dispose de 256 niveaux de gris dont 40 sont reconnus à l'œil nue. Plus le nombre de bit est grand plus les niveaux sont nombreux et plus la représentation est fidèle. [4]



Figure I.4 : Image en niveau de gray.

I.5.5 Image couleur (RGB)

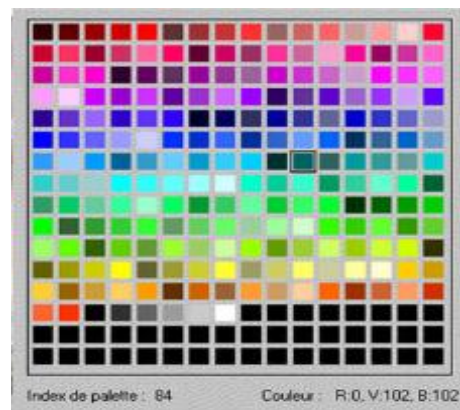
Chaque pixel possède une couleur décrite par la quantité de rouge (R), vert (G) et bleu (B). Chacune de ces trois composantes est codée sur l'intervalle $[0, 255]$, ce qui donne $255^3 = 16\,777\,216$ couleurs possibles. Il faut 24 bits pour coder un pixel.

Codage d'une image en couleurs 8 bits

Dans ce cas on attache une palette de 256 couleurs à l'image. Ces 256 couleurs sont choisies parmi les 16 millions de couleurs de la palette RVB. Pour chaque image le programme recherche les 256 couleurs les plus pertinentes. Chaque code (de 0 à 255) désigne une couleur. L'image occupe 3 fois moins de place en mémoire qu'avec un codage 24 bits. L'image est moins nuancée : sa qualité est bonne mais moindre.



Image en couleurs 8 bits



Palette 8 bits

Figure I.5 : Représentée l'image en couleurs 8 bits et la palette 8 bits. [13]

Image couleur 8 bits et sa palette. Remarquer la couleur 84 et sa correspondance en RVB.

Codage d'une image en couleurs 24 bits

Il existe plusieurs modes de codage de la couleur. Le plus utilisé est le codage Rouge, Vert, Bleu (RVB). Chaque couleur est codée sur 1 octet = 8 bits. Chaque pixel sur 3 octets c'est à dire 24 bits : le rouge de 0 à 255, le vert de 0 à 255, le Bleu de 0 à 255.

Le principe repose sur la synthèse additive des couleurs : on peut obtenir une couleur quelconque par addition de ces 3 couleurs primaires en proportions convenables.

On obtient ainsi $256 \times 256 \times 256 = 16777216$ (plus de 16 millions de couleurs différentes).

Table I.1 – Tableau du codage d'une image en couleurs 24 bits

Couleur	Noir	Bleu pale	Vert pale	Rose	Bleu	Vert	Rouge	Blanc
R	0	120	120	150	120	120	255	255
V	0	120	150	120	120	255	120	255
B	0	150	120	120	255	120	120	255

L'image de 10000 pixels ainsi codée occupe $10000 \times 3 = 30000$ octets.

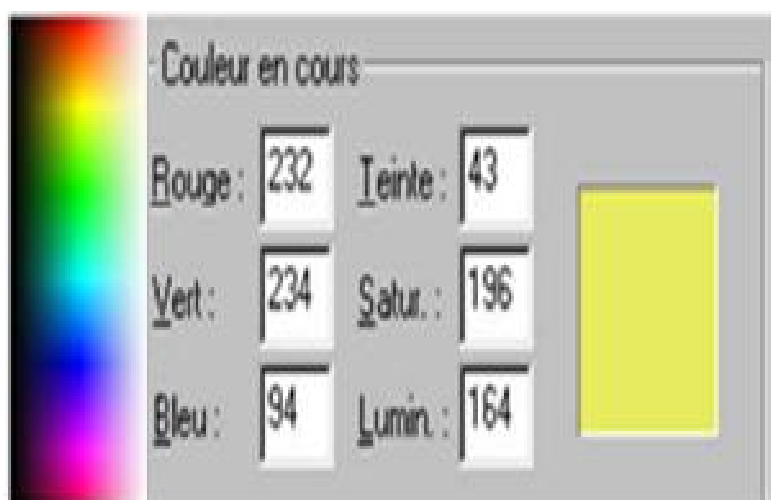


Figure I.6 – Palette 24 bits Exemple de couleur en 24 bits



Figure I.7 – Image en couleurs 24 bits

Dans le cas d'une image de 10 cm x 10 cm avec une résolution convenable de 100 pixels par cm (un pixel mesure 0,1 mm), elle est codée sur $1000 \times 1000 = 1\,000\,000$ pixels

Elle occupe : en noir et blanc : $1\,000\,000$ bits = $125\,000$ octets

En 256 niveaux de gris : 1 million d'octets

En couleurs (24bits) : 3 millions d'octets.

C'est codage de la couleur qui est utilisé par la plupart des écrans d'ordinateurs actuellement.

On constate qu'il est très gourmand en mémoire. Pour faciliter le stockage des images en mémoire on utilise d'autres formes de codage.

I.5.6. Contraste

C'est l'opposition marquée entre deux régions d'une image. Une image contrastée présente une bonne dynamique de la distribution des valeurs de gris sur tout l'intervalle des valeurs possibles, avec des blancs bien clairs et des noirs profonds. Au contraire une image peu contrastée a une faible dynamique, la plupart des pixels ayant des valeurs de gris très proches.

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

$$C = \frac{L1-L2}{L1+L2}$$

I.5.7 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.

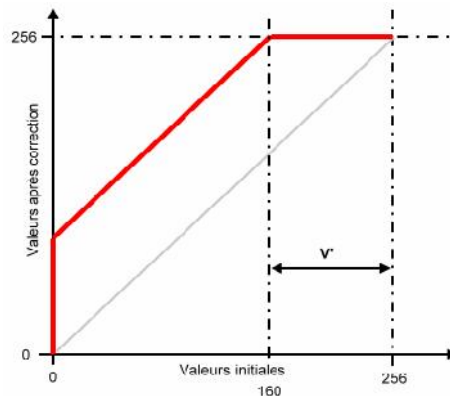


Figure I.8 : Pour une valeur de +96, tous les points de l'espace V' seront blancs.

Une bonne luminance se caractérise par :

- Des images lumineuses (brillantes).
- 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.
- L'absence de parasites.



Figure I.9 : Changer la luminance dans l'image (a) et (b).

I.5.8. 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 [Gonzales & Wintz, 77]. C'est un parasite qui représente certains défauts (poussière, petits nuages, baisse momentanée de l'intensité électrique sur les capteurs, ...etc.). Il se traduit par des taches de faible dimension et dont la distribution sur l'image est aléatoire.

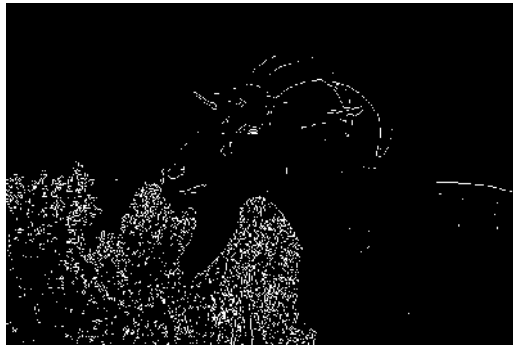


Figure I.10 : Exemple sur l'image sans bruit et bruitée. [5]

I.5.9 Contour

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentant une différence significative.

Dans une image numérique, les contours se situent entre les pixels appartenant à des régions ayant des intensités moyennes différentes; il s'agit de contours de type « saut d'amplitude ». Un contour peut également correspondre à une variation locale d'intensité présentant un maximum ou un minimum; il s'agit alors de contour « en toit ».



(a) Contour détectés par le filtre de Sobel.



(b) Contour détectés par le filtre de Canny.

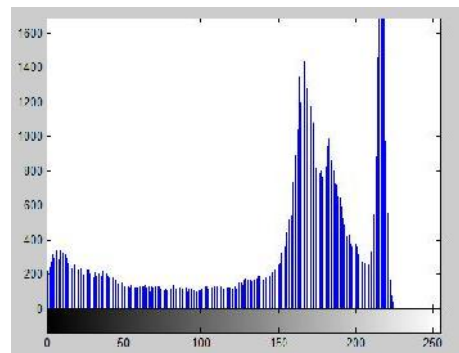
Figure I.11 : Détection de contour sur l'image Chever.

I.5.10 Histogramme de l'image. [5]

L'histogramme est un vecteur dont la dimension est égale à 256 (dans le cas d'une image à niveau de gris). Chaque élément de ce vecteur (noté $h(i)$) représente le nombre de pixels de l'image possédant le niveau de gris i (sa fréquence d'apparition), on peut donc assimiler l'histogramme à la densité de probabilité des intensités lumineuses.

Puisque le nombre des pixels est généralement assez grand, on peut alors normaliser l'effectif de chaque niveau de gris en divisant chaque terme du tableau par la surface du plan exprimée en nombre total des pixels de l'image. De ce fait, chaque case $h(i)$ du vecteur représente la probabilité d'avoir l'intensité i .

Il peut être utilisé pour améliorer la qualité d'image (rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci.

**Figure I.12** : Image avec histogramme.

I.6. Types d'image

I.6.1 Images matricielles (ou image bitmap)

Elle est composée d'une matrice (tableau) de points à plusieurs dimensions, chaque dimension représentant une dimension spatiale (hauteur, largeur, profondeur), temporelle (durée) ou autre (par exemple, un niveau de résolution).



Figure I.13 : Exemple d'une image matriciel. [6]

I.6.1.1 Images 2D

Dans le cas des images à deux dimensions (le plus courant), les points sont appelés pixels. D'un point de vue mathématique, on considère l'image comme une fonction de $\mathbb{R} \times \mathbb{R}$ dans \mathbb{R} où le couplet d'entrée est considéré comme une position spatiale, le singleton de sortie comme un codage.

Ce type d'image s'adapte bien à l'affichage sur écran informatique (lui aussi orienté pixel); il est en revanche peu adapté pour l'impression, car la résolution des écrans informatiques, généralement de 72 à 96 ppp (« points par pouce », en anglais dots per inch ou dpi) est bien inférieure à celle atteinte par les imprimantes, au moins 600 ppp aujourd'hui. L'image imprimée, si elle n'a pas une haute résolution, sera donc plus ou moins floue ou laissera apparaître des pixels carrés visibles.

I.6.1.2. Image 2D +t (vidéo), image 3D, image multi-résolution

- Lorsqu'une image possède une composante temporelle, on parle d'animation.
- Dans le cas des images à trois dimensions, les points sont appelés des «voxel ». Ils représentent un volume.

Ces cas sont une généralisation du cas 2D, la dimension supplémentaire représentant respectivement le temps, une dimension spatiale ou une échelle de résolution. D'un point de vue mathématique, il s'agit d'une fonction de $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ dans \mathbb{R} .

I.6.1.3 Images stéréoscopique

Il s'agit d'un cas particulier dans lequel on travaille par couples d'images, ces derniers pouvant être de n'importe lequel des types précédents.

Il existe un grand nombre de sortes d'images stéréoscopiques, et un encore plus grand nombre de moyens pour les observer en relief, mais le codage recommandé par les organisations internationales de stéréoscopie est désigné comme « jps », c'est-à-dire un format JPG dans lequel les deux vues gauche et droite sont juxtaposées dans un même fichier, le plus souvent $2\,048 \times 768$, chacune des deux vues étant inscrite dans un rectangle $1\,024 \times 768$ et, si son rapport largeur/hauteur n'est pas $4/3$, chaque vue est complétée dans ce rectangle par deux bandes noires symétriques, soit en haut et en bas, soit à gauche et à droite.

I.6.2 Images vectorielles

Le principe est de représenter les données de l'image par des formules géométriques qui vont pouvoir être décrites d'un point de vue mathématique. Cela signifie qu'au lieu de mémoriser une mosaïque de points élémentaires, on stocke la succession d'opérations conduisant au tracé. Par exemple, un dessin peut être mémorisé par l'ordinateur comme « une droite tracée entre les points (x_1, y_1) et (x_2, y_2) », puis « un cercle tracé de centre (x_3, y_3) et de rayon 30 de couleur rouge ».

I.7 Traitement d'image

I.7.1 Définition de Traitement

C'est un ensemble des opérations relatives à la collecte, à l'enregistrement, à l'élaboration, à la modification, à l'édition, ... données. Mettons de cotés les termes enregistrement et édition. Le principe général du traitement d'image est donc à quelques détails près toujours le même un système reçoit des images, y applique un traitement, et produit une information de nature liée à l'application visée.

Une source de rayonnement envoie des ondes sur un objet, qui sont ensuite réfléchies et collectées par un capteur. Le capteur transforme ces ondes en un ensemble de points. Ces points sont traités et une information est produite en sortie du système. On peut résumer le traitement d'image en quatre étapes principales :

➤ Acquisition des images :

Mise en œuvre des processus physiques de formation des images suivis d'une mise en forme pour que ces images puissent être traitées par des systèmes informatiques.

➤ Traitement des images :

Son but : améliorer ces images lorsqu'elles possèdent du bruit ou des défauts.

➤ Segmentation des images :

Son but : construire une image symbolique en générant des régions homogènes selon un critère défini à priori.

➤ Analyse des images :

Consiste à extraire des paramètres ou des fonctions représentatives de l'image ou des Régions.

I.7.2 les techniques fondamentales en traitement d'images

Une panoplie de traitements peut être appliquée à l'image numérique, la figure suivante résume l'ensemble de ces traitements.

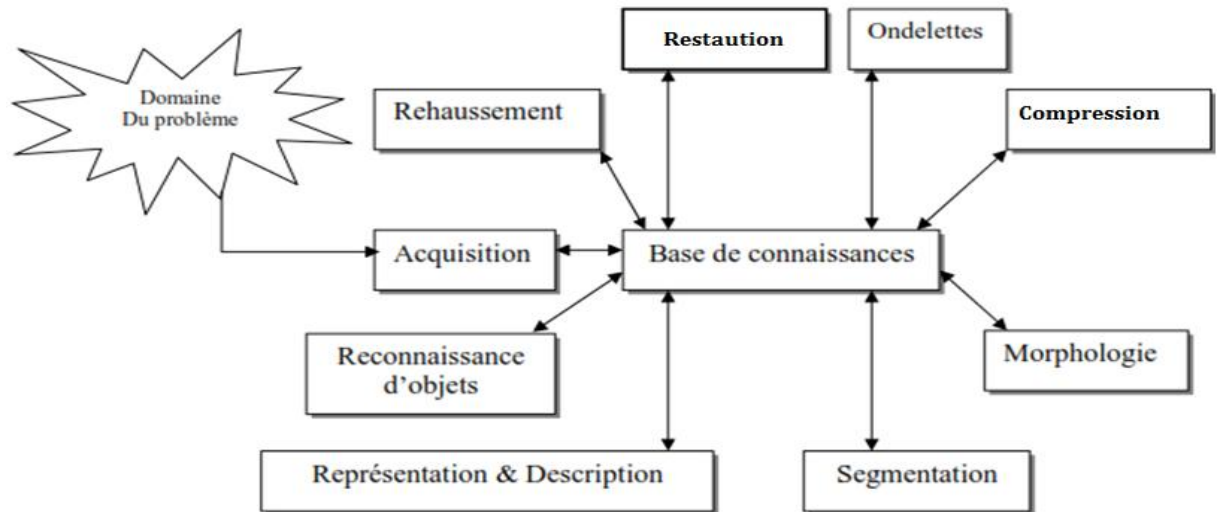


Figure I.15 : Traitements fondamentaux en traitement d'images.

I.7.2.1 Acquisition de l'image

L'acquisition d'images est une mesure spatiale d'une interaction entre une onde et de la matière. L'onde est émise par une source et reçue par un capteur. La matière occupe de l'espace et possède une masse. Elle a pour objet de passer de la scène physique à une forme numérique observée.

I.7.2 Le rehaussement de l'image

Le mécanisme de formation des images est loin d'être parfait donc présence de différentes formes de bruit, améliorer le contraste d'où l'objectif est de :

- Rehausser le niveau de gris
- Accentuer les caractéristiques

I.7.2.3 Les ondelettes

Elles permettent de représenter l'image en différentes résolutions. Elles sont utilisées aussi en compression.

I.7.2.4 La morphologie

C'est un outil permettant d'extraire des composantes d'une image pour décrire et représenter différentes formes.

I.7.2.5 La segmentation

C'est une procédure permettant de partitionner l'image en ses constituants ou objets. La segmentation automatique est la tâche la plus difficile en traitement d'images. Plus la segmentation est meilleure plus l'étape de reconnaissance d'objets est réussie.

I.7.2.6 La représentation et description

La représentation est l'étape qui vient juste après la segmentation. Le résultat de la segmentation est un ensemble de pixels relatifs à une région. Ces données doivent être converties en une forme traitable par un ordinateur. On peut soit représenter la région ou sa frontière. La description est l'extraction d'attribut permettant de distinguer une classe d'objets d'une autre classe.

I.7.7 La reconnaissance : [8]

Est le traitement qui affecte une étiquette (exemple : route, voiture,..) à un objet en se basant sur ses descripteurs.

La base de connaissance contient la connaissance du domaine du problème en cours du traitement. Dans son aspect le plus simple, elle peut consister en coordonnées de l'objet à traiter, ceci permet de réduire l'espace de recherche. Comme elle peut être complexe contenant toutes les défaillances que peut présenter un produit manufacturé.

I.7.2.8 La restauration d'image

Les images subissent des dégradations, ces dégradations sont dues, d'une part aux Appareils d'acquisition et d'autre part aux conditions de prise de vue...etc. Son but est d'améliorer la qualité d'une image, atténuer, supprimer les dégradations

La restauration se fait par des techniques comme noir et blanc, négative, rotation, ...etc.

Noir et blanc : L'image noir et blanc, est une image dont les couleurs ont été remplacés par le gris sauf le noir et le blanc de l'image.

Négative : est une image dont les couleurs ont été inversées par rapport à l'originale par exemple le noir devient blanc et inversement.

Compressions d'image : La compression d'image est une application de la compression de données sur des images numériques. Cette compression a pour utilité de réduire la redondance des données d'une image afin de pouvoir l'emmagasiner sans occuper beaucoup d'espace ou la transmettre rapidement.

La compression d'image peut être effectuée avec perte de données ou sans perte

Compression sans perte : Appelée aussi compression non destructrice, la qualité de l'image après décompression est la même que celle de l'image originale, le taux de compression de ce type est limité.

Ce type de compression on le trouve beaucoup dans le domaine où la précision est majeure. Comme l'image médicale (IRM par ex.) ou la télédétection (imagerie satellite par ex.).

Compression avec perte : C'est une compression destructrice, elle permet de sacrifier certains détails de l'image non récupérable en décompression au profit de réduction de poids. Cette dégradation peut être contrôlée selon la qualité qu'on veut obtenir en fonction du taux de compression choisie.

Ce type de compression on le trouve généralement dans le domaine normal pratique là où la réduction du poids de l'images est très important, comme le domaine multimédia par exemple (web, photographie) où la fidélité envers l'image original n'est pas très importante et le taux de compression sera plus grand que celui d'une compression sans perte du fait qu'on est juste limité par la qualité qu'on souhaite obtenir.

I.7.3 Historique

Le traitement d'image commence à être étudié vers les années 1920, pour la transmission d'images par le câble sous-marin allant de New York à Londres. En effet, la

première numérisation d'image avec compression de données pour envoyer des fax de Londres à New York¹.

Les années 1950, les premiers travaux sur l'analyse d'images dans les chambres à bulles. Les années 1960, ont vu le véritable départ du traitement d'images quand les ordinateurs commencent à être suffisamment puissants pour travailler sur des images.

Les premiers essais sur l'extraction automatique d'information ont commencé avec le début de la décennie 70. Après les années 80, il y a eu un passage de l'image 2D aux modèles 3D.

I.8 Conclusion

Dans ce chapitre, nous avons présenté les différents concepts liés à l'image numérique, on a parlé sur les types d'une image, ces caractéristiques et quelques formats de fichiers d'images, on a expliqué aussi la structure d'un fichier bitmap et finalement nous avons présenté quelques traitements d'images usuels.

Le traitement d'images est l'ensemble des méthodes et techniques opérant sur celles-ci, dont le but est de rendre cette opération possible, plus simple, plus efficace et plus agréable, d'améliorer l'aspect visuel de l'image et d'en extraire des informations utiles. [9]

Chapitre II

Segmentation d'images

II.1 Introduction

Une image est constituée d'un ensemble de pixels.

Segmenter une image en régions, c'est rassembler ces pixels en ensemble ayant des propriétés communes, comme une couleur ou une texture similaire, ou un mouvement cohérent.

Segmenter une image en objets, c'est rassemble des régions qui ont un sens sémantique commun, même si leurs propriétés sont différentes. Il s'agit par exemple de distinguer un personnage du contexte dans lequel il évolue, alors que sette personne a des vêtements de couleurs différentes ou ne bouge que certaines parties de son corps.

L'importance de la segmentation croît avec celle de l'image dans notre société. En imagerie médicale, la segmentation peut aider le médecin dans son diagnostic. En compression vidéo, elle permet de traiter différemment une zone d'intérêt, qui bénéficiera d'une plus grande précision, du reste de l'image qui pourra être plus fortement compressé. En indexation, la segmentation sert à extraire un objet que l'on souhaiterait retrouver dans d'autres images. En postproduction cinématographique, un personnage segmenté peut être replace dans un autre décor, ce qui le cas notamment pour les présentateurs de la météo. En vidéosurveillance, la détection d'objets en mouvement peut révéler la présence d'intrus, ou évaluer la fluidité d'un trafic routier. Les applications sont nombreuses et la liste des exemples cités est loin d'être exhaustive.

Mais revenons au principe de la segmentation. Pour segmenter une région, une méthode consiste à détecter son contour ?le passage d'une région à une autre se caractérise par une différence d'intensité lumineuse. Plus la différence est importante, plus le gradient de l'image sera fort, et plus le contour sera marqué. Prenons pour exemple l'image de la **figure I.1** et essayons de segmenter la coccinelle.



Figure II.1 – Segmentation : image originale.

La figure 1.2 nous présente deux extractions possible des gradients. Ces gradients définissent des contours qui n'ont pas tout le même intérêt pour nous. Les contours de la feuille sont bien marqués, alors que le contour de la coccinelle n'a pas une intensité constante.

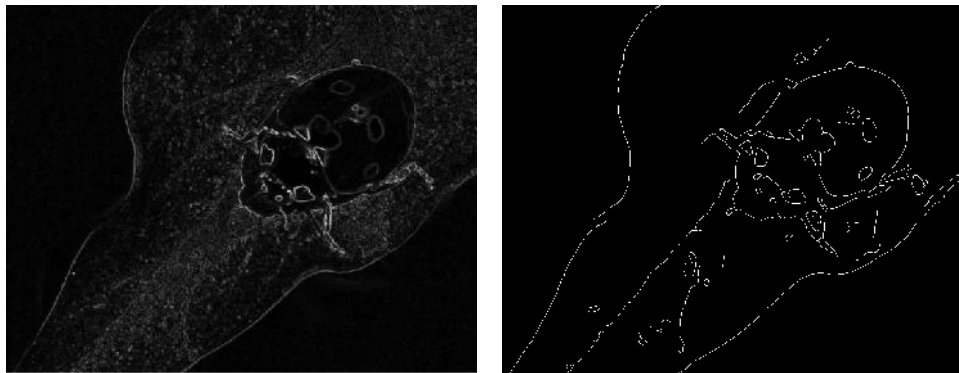


Figure II.2 -Extraction des contours d'une image à partir de ses gradients.

Si l'on peut déterminer l'intensité à laquelle les gradients sont détectés, on ne peut cependant pas déterminer leur nombre, s'assurer de leur connexité, ou leur donner un sens quant à leur appartenance à l'objet. Autrement dit, tout les gradients détectés n'appartiennent pas au contour de la coccinelle, et l'on ne saurait les faire disparaître sans perdre de la même manière une partie des contours de la coccinelle. De plus, ces contours ne sont pas forcément fermés ne définissent donc pas des régions séparées. [10]

Certains algorithmes permettent de définir des régions à partir des gradients de l'image. Là encore, on peut contrôler le nombre de régions qui découpent l'image, mais on ne peut savoir si ces régions appartiennent à l'objet d'intérêt ou non. La figure 1.3, la coccinelle est bien détournée, mais certaines régions de la coccinelle (les régions blanches) sont considérées comme n'appartenant pas à la coccinelle. D'autres régions sont segmentées, mais elles ne nous intéressent pas.



Figure II.3 : Segmentation en régions à des contours morphologiques par watershed. [3]

Une autre approche consiste à définir directement les régions que l'on souhaite segmenter, à partir d'une caractéristique commune. Par exemple,



Figure II.4 – Segmentation en régions de couleur homogène.

On peut choisir de segmenter l'image en régions de même couleur, ou plus précisément de couleur homogène (voir figure I.4). On rencontre le même problème que précédemment, à savoir que notre coccinelle est constituée de plusieurs régions de différentes couleurs et qu'on ne l'a toujours pas distinguée du reste de l'image.

La dernière approche que nous verrons consiste à définir un contour fermé et à la faire évoluer vers l'objet d'intérêt (voir la figure I.5). La tâche est difficile et nécessite d'avoir une idée de l'objet ou de la région que l'on souhaite segmenter. Pour définir cet objet, il faut choisir un critère qui détermine si un pixel fait partie de l'objet ou non. Ce critère qui la description des propriétés de l'objet. Le contour actif peut par exemple évoluer vers des zones de fort gradient, tout en essayent d'entourer région la plus homogène possible au niveau de la couleur.

De nombreuses propriétés peuvent être prises en compte dans le critère, et le contour actif évolue en effectuant un compromis entre elles. En cas d'objets multiples, il peut également se séparer ; au contraire, deux contours peuvent fusionner.



Figure II.5 : Segmentation par contour actifs : du contour initial (à gauche) au contour à convergence (à droite).

II.2. Définition de la segmentation d'image

La segmentation d'image est un traitement de bas niveau qui consiste à créer une partition de l'image I en termes d'un prédicat d'homogénéité P et d'un ensemble de régions R_i vérifiant les critères suivants :

- $I = \cup_i R_i$
- $R_i \neq \emptyset \forall_i$
- $R_i \cap R_j = \emptyset, \forall i, j; i \neq j$
- R_i est connexe $\forall i$
- $P(R_i) = \forall i$
- $P(R_i \cup R_j)$ faux $\forall i \neq j$ et R_i, R_j sont adjacent

Donc une région est un ensemble connexe de points d'image appelés pixels ayant des propriétés communes tel que l'intensité, la texture, etc. qui les différencient des pixels des régions voisines. [11]

Il n'y a pas de méthode unique de segmentation d'une image, le choix d'une technique est lié :

- à la nature de l'image :
 - éclairage non homogène, reflets,
 - présence de bruit, de zones texturées,

- contours flous, en partie occultés,
- ✚ aux opérations situées en aval de la segmentation :
 - localisation, mesure, calcul 3D,
 - reconnaissance des formes, interprétation,
 - diagnostic, contrôle qualité,
- ✚ aux primitives à extraire :
 - contours, segments de droite, angles,...
 - régions, formes,
 - textures,
- ✚ aux contraintes d'exploitation :
 - complexité algorithmique, fonctionnement en temps réel,
 - taille de la mémoire disponible en machine.

II.3. Les différentes approches de segmentation

Essentiellement, l'analyse d'images a pour but l'extraction de l'information caractéristique contenue dans une image. Cette information peut prendre la forme, la couleur, le contour, la texture,... Donc, il est nécessaire de procéder tout d'abord à la segmentation de la lésion. Cette segmentation est généralement abordée selon deux modèles complémentaires : l'approche région et l'approche frontière.

L'approche frontière regroupe les techniques associées à une variation d'intensité : détection de contours. Cependant, les contours obtenus sont rarement connexes, donc elle doit faire face au problème de fermeture de contours.

L'approche région considère des groupements de pixels ayant des propriétés communes, chaque pixel est affecté à une région unique après une partition de l'image.

Toutefois, il n'existe pas d'algorithme spécifique pour déterminer les régions et les frontières dans une image. D'ailleurs les chercheurs ont compromis que la segmentation idéale n'existe pas, la bonne technique de segmentation sera donc celle qui permettra d'arriver à une bonne interprétation. [12]

Une panoplie de techniques est proposée dans la littérature où chacune présente des avantages et des inconvénients. Parmi les nombreuses approches de segmentation d'images qui existent, nous présentons quelques techniques les plus citées dans la littérature appliquées à la mammographie, à savoir les méthodes basées sur pixels, l'approche région ainsi que celles basées sur les contours.

II.3.1. Segmentation basée sur les contours

La détection de contour est souvent le premier problème qu'on rencontre en traitant une image. La difficulté augmente avec l'importance de bruit présent.

II.3.1.1 Définition (un contour) :

Un contour peut approximativement être défini comme une frontière entre deux régions ou l'intensité des pixels change brusquement.

Généralement l'utilisation d'un tel opérateur de contour se combine avec un seuillage et comme étant ce dernier est généralement imparfait, on obtient, d'une part, des contours qui ne limitent pas les régions fermées. Donc on doit faire recours à des algorithmes de fermetures des contours. D'un autre part, les zones de fortes variations ne correspondent pas forcément à un contour d'objet. Alors un post-traitement est nécessaire pour analyser les différents contours obtenus. [13]

II.3.1.2 Méthode des contours actifs (snacke)

L'idée de cette méthode est de déplacer les points pour les rapprocher des zones de fort gradient tout en conservant des caractéristiques comme la courbure du contour ou la répartition des points sur le contour ou d'autres contraintes liées à la disposition des points. Au démarrage de l'algorithme, le contour est disposé uniformément autour de l'objet à détecter puis il va se rétracter pour en épouser au mieux la forme. À chaque itération, l'algorithme va tenter de trouver un meilleur positionnement pour le contour pour minimiser les dérives par rapport aux contraintes utilisées. L'algorithme s'arrêtera lorsqu'il ne sera plus possible d'améliorer le positionnement ou simplement quand le nombre maximum d'itérations aura été atteint. On utilise les notions d'énergies interne et externe pour caractériser respectivement la forme du contour et son positionnement sur l'image en tenant compte des lignes de gradient.

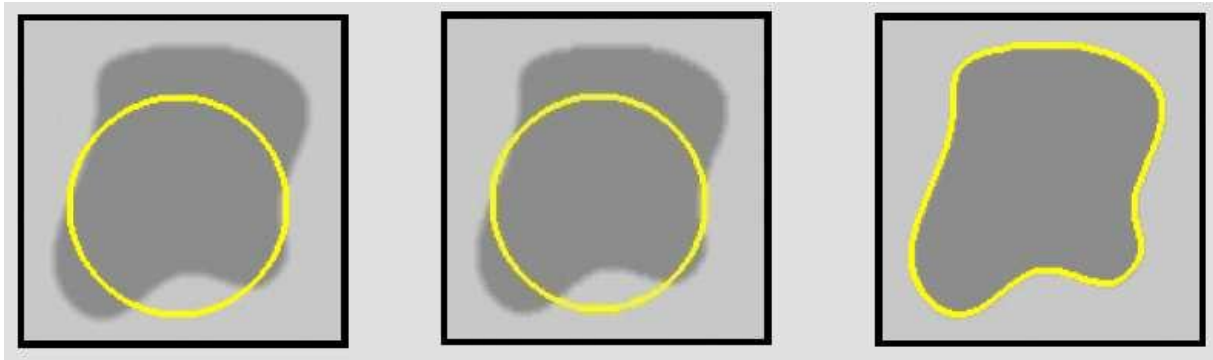


Figure II.6 : principe de contour actif.

II.3.1.3 Méthode d'ensembles de niveau zéro (LEVEL SETS)

La méthode des Ensembles de Niveau Zéro est une méthode de simulation numérique utilisée pour l'évolution des courbes et des surfaces dans les domaines discrets.

L'idée de base de la méthode des level sets est de considérer une courbe (ou interface) en mouvement comme le niveau zéro d'une fonction de dimension plus élevée. Pour une courbe en 2D, cette interface () est l'intersection d'une hyper surface (de dimension 3) avec un plan. [14]

Les points définissant cette interface vont se déplacer vers la normale à une vitesse F selon

L'équation suivante :

$$F \frac{\partial \phi}{\partial t} + \nabla \phi \cdot \mathbf{V} = 0 \quad (1)$$

Cette vitesse F est composée de trois termes : un terme constant (similaire à la force d'inflation utilisée dans les modèles déformables), un terme dépendant de la courbure locale

en chaque point et un terme dépendant de l'image (dans notre cas, les fronts de l'image). Le schéma numérique de l'équation de déplacement de l'interface est décrit par l'équation :

$$x_{n+1} = x_n - dt * k_1(x, y) * (U_n \cdot \varepsilon k) * \nabla_t \varepsilon [0,1] \quad (2)$$

Avec :

Un $(m, \sigma) = \pm 1$, fonction d'appartenance définissent la zone ou l'objet à recherche.

$$K = \nabla \cdot \frac{\nabla \Psi}{|\nabla \Psi|}, \text{ courbure locale en chaque point de l'interface.} \quad (3)$$

$KI(x, y)$, critère d'arrêt dépende de l'image de gradient. L'initialisation est réalisée avec une ou plusieurs de départ.

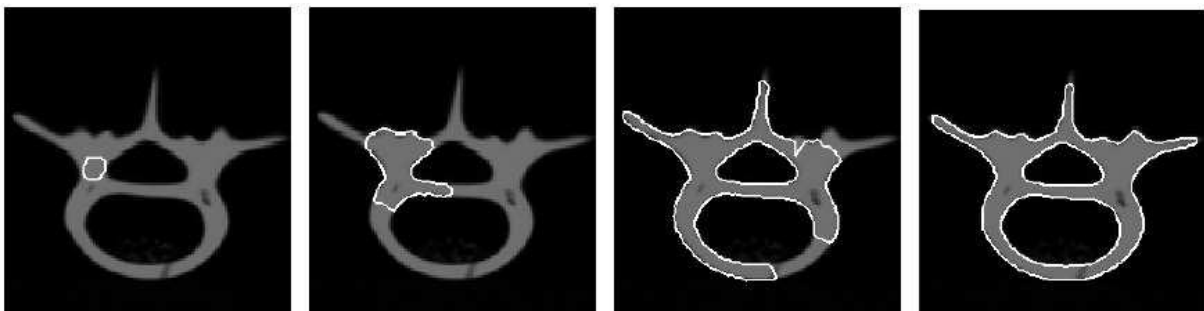


Figure II.7 : principe de Level Sets.

L'avantage principal de cette méthode est la possibilité de gérer automatiquement le changement de topologie de la courbe en évolution.

Et ses limitations :

- Problème de divergence : Méthode de résolutions.
- Interpolation : perte de matière.
- Cout de calcul : Réinitialisation de la fonction de distance. [15]

II.3.2. Approches de segmentation par région

L'approche par régions consiste à regrouper des points selon des propriétés communes.

II.3.2.1. Définition (Segmentation par région)

Globalement, elle peut être définie comme une partition d'une image I en une ou plusieurs régions R_1, \dots etc. R_n telles que :

$$I = \bigcup_{i=1}^n R_i \text{ et } R_i \cap R_j = \emptyset \text{ pour } i \neq j \quad (4)$$

II.3.2.1.1 Méthodes par croissance de régions

Le principe de ces méthodes est de réunir de façon itérative un ensemble de points connectés en une région de plus en plus large, en fonction de critères d'homogénéité. Pour définir une région, on définit un germe dans la région d'intérêt qui sert comme un point de départ pour l'agrégation (Figure II.4).

- ❖ Cependant un inconvénient à marquer à cette méthode c'est qu'elle est récursive: risque de débordements (pile).

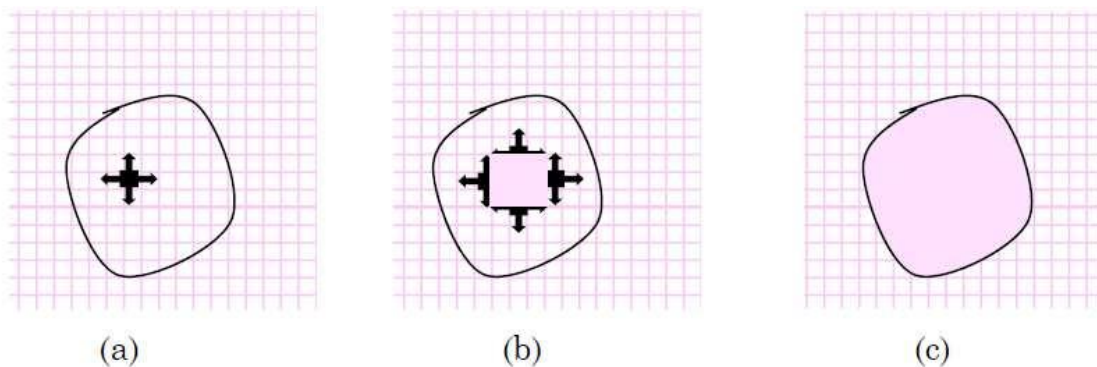


Figure II.8 : Le processus d'un algorithme de croissance de région. (a) début de processus, le point noir présente le germe et les flèches la direction de croissance. (b) la croissance de pixels après quelques itérations. (c) le résultat de la segmentation.

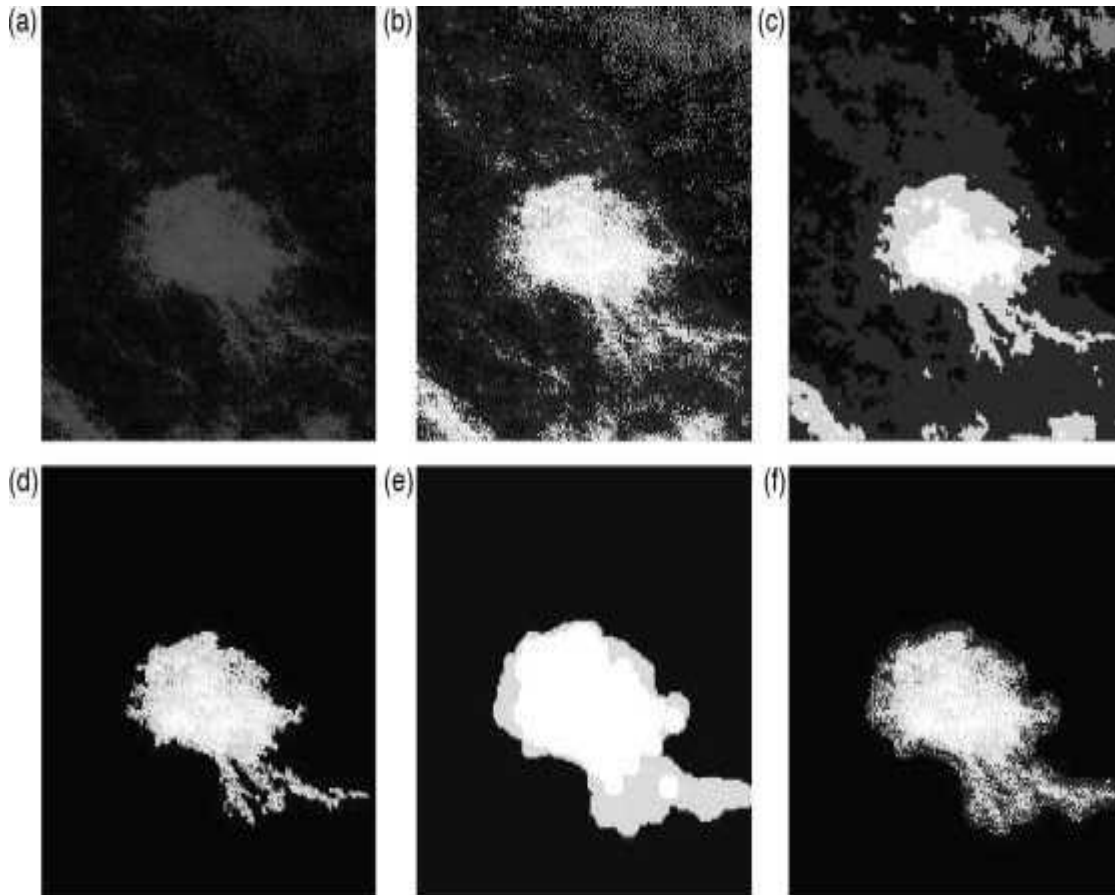


Figure II.9 : L'application de l'algorithme croissance de région sur les images mammographiques.

❖ Segmentation morphologique : LPE

La Ligne de Partage des Eaux (LPE) est une technique de segmentation d'image assez répandue. Cet outil consiste à chercher les contours des bassins versants dans l'image à segmenter. Après avoir déterminé les contours de ces bassins qui forment les minima locaux, on obtient une carte de régions. L'inconvénient majeur de cette technique réside dans sa sensibilité au nombre de germes (généralement les minima locaux) qui s'avère, souvent, très important et engendre donc une sur-segmentation de l'image. [16]

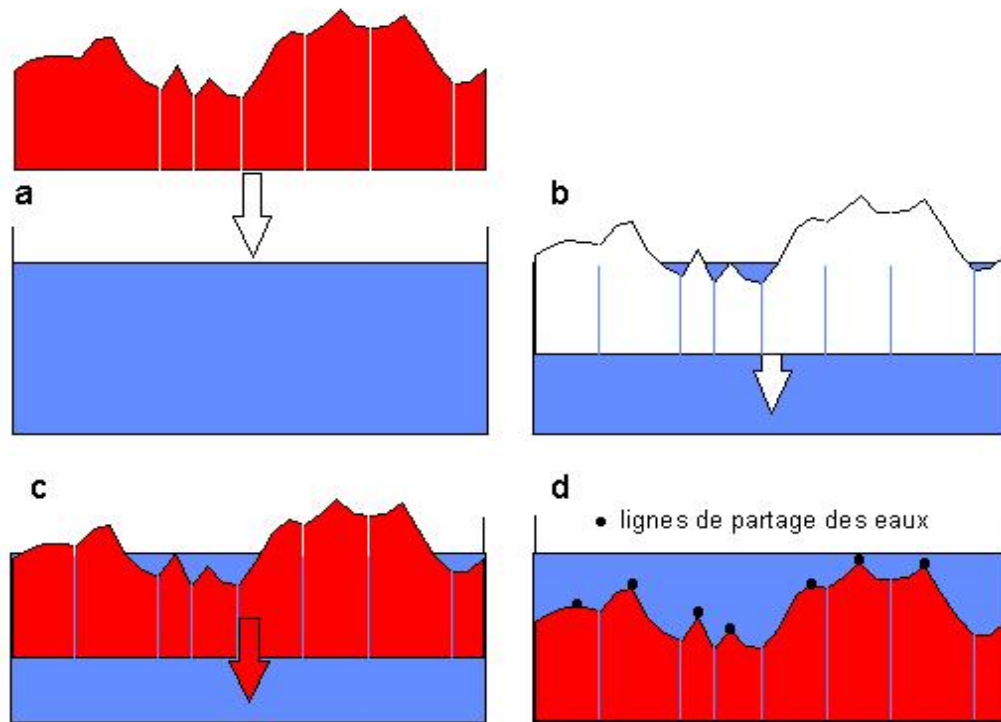


Figure II.10 : Principe de la Ligne de Partage des Eaux.

❖ **K-means**

Cette méthode consiste à rassembler les pixels en K groupes (clusters). K étant un paramètre préfixé qui détermine le nombre de régions. Et après avoir déterminé les paramètres (couleur par exemple) de toute région, chaque point est affecté au centre le plus proche, pour recalculer ensuite le paramètre de chaque région, jusqu'à ce qu'elles soient stables.

✓ **Algorithme K-means :**

L'algorithme des K-Means est l'une des techniques de clustering non supervisée les plus utilisées. La méthode consiste à placer aléatoirement dans l'espace K "cent roide" afin de déterminer K clusters. On affecte à chacun de ces "cent roides" les objets les plus proches, puis on calcule la position moyenne des objets associés aux "cent roide" que l'on déplace en ce point. Les opérations d'affectation d'objet et de déplacement du cent roide sur la moyenne répétées jusqu'à ce que chaque cent roide ait atteint une position stable.

Malgré sa simplicité, cet algorithme se révèle efficace. Toutefois il est nécessaire de prédéterminer le nombre de catégories et la position de départ des prototypes qui a un impact sur le découpage en classes. Les principales étapes de cet algorithme sont :

- Choix aléatoire de la position initiale des K clusters.
- (Ré-)Affecter les objets à un cluster suivant un critère de minimisation des distances (généralement selon une mesure de distance euclidienne). [17]
- Une fois tous les objets placés, recalculer les K centres.
- Répéter les étapes 2 et 3 jusqu'à ce que plus aucune réaffectation ne soit faite.



Image original

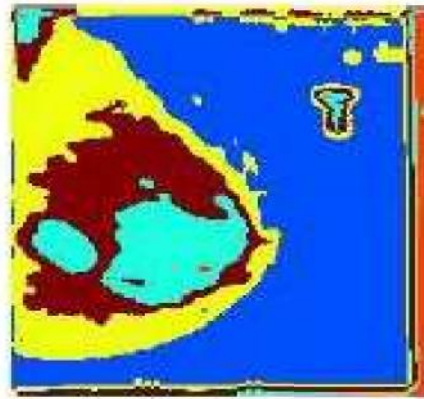


Image segmentée avec $K=5$
(K est le nombre de classe)

Figure II.11 : L'application de l'algorithme K-means.

Le principal avantage de cette méthode est que le nombre de régions est connu au préalable (il n'y aura pas de problème de sur-segmentation ou sous-segmentation), mais l'inconvénient réside en la difficulté de déterminer le nombre de clusters et l'incohérence des régions (deux objets éloignés peuvent appartenir à la même région).

II.3.3. Segmentation basée sur les pixels

La segmentation basée sur le pixel travaille sur des histogrammes de l'image, on citera essentiellement la méthode du seuillage. La segmentation est inhérente puisque les régions sont déduites du seuillage.

II.3.3.1 Seuillage de l'image

Le seuillage est une étape fondamentale pour détecter les contours significatifs. Il est basé sur l'hypothèse que le bruit présent dans l'image est relativement faible par rapport au signal. L'opération dite de seuillage simple consiste à mettre à zéro tous les pixels ayant un niveau de gris inférieur à une certaine valeur et à la valeur maximale les pixels ayant une valeur supérieure. Ainsi le résultat du seuillage est une image binaire contenant des pixels noirs et blancs. En traitement d'images, on préfère souvent au seuillage simple, un seuillage par hystérésis utilisant deux valeurs de seuils. Le seuillage par hystérésis repose sur une hypothèse de connexité. Il permet d'éliminer les points de contours trop faibles et de ne conserver que les composantes connexes les plus importantes des contours. Le principe est d'utiliser deux seuils pour la norme du gradient : s_h et s_b . Les pixels dont la norme du gradient est supérieure à s_b sont sélectionnés. Si ces pixels sont connectés, par un chemin de pixels dont la norme est supérieure à s_b , à un pixel de la norme est supérieure à s_h , alors le contour est retenu. [18]

Le choix des valeurs des seuils peut être automatisé, mais il est le plus souvent guidé par le jugement visuel. La Figure II.12(a) représente les contours obtenus sur l'image de Léna avec un filtre de deriche. Les valeurs des seuils ont été fixées de manière à obtenir le meilleur rapport entre les contours désirés et les contours non-désirés. Si les valeurs seuil choisies sont trop faibles, l'image contient un grand nombre de contours non-désirés (Figure II.12(b)). Il s'agit en réalité de maxima locaux de faibles valeurs liés au bruit qui ont été assimilés à des contours. On obtient alors de faux-contours qui n'existent pas dans la scène réelle. Au contraire, lorsque les valeurs seuil sont trop importantes, certains contours présents dans l'image sont

supprimés de l'image (Figure II.12(c)). Plus le seuil est élevé, plus le risque de perdre les contours de l'image est important. De plus, pour des valeurs de seuils proches, il est souvent difficile de déterminer la meilleure détection (Figure II.12(d)). La valeur de seuil des images est donc relativement difficile à estimer.

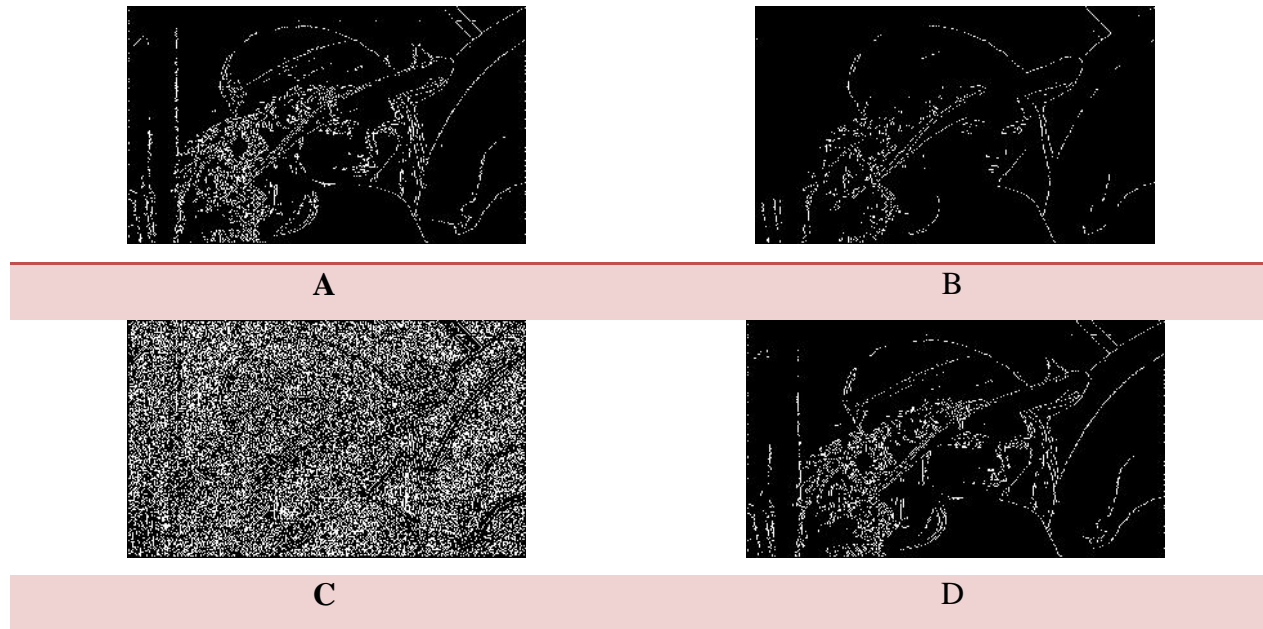


Figure. II.12 : Detections de contours sur l'image de Léna pour différentes valeurs de seuils (a) $s_h = 29$ et $s_b = 26$, (b) $s_h = 0$ et $s_b = 0$, (c) $s_h = 60$ et $s_b = 55$, (d) $s_h = 32$ et $s_b = 30$.

Supposons que nous ayons réussi à trouver les meilleures valeurs de seuils pour l'image de Lena. Nous souhaitons à présent traiter d'autres images. Ces valeurs sont-elles des valeurs de seuils universelles ? Et si non, existe-t-il des valeurs de seuils universelles qui fonctionnent pour toutes les images ?

Considérons une seconde image, celle des poivrons représentée Figure II.13(a). Cette image est traitée comme l'image de Lena, avec un filtre de Deriche. La détection de contours est effectuée avec les mêmes seuils que l'image de Lena (Figure II.12(a)). Sur cette image, nous souhaiterions détecter des contours fermés pour chaque poivron et piments. Après traitement, l'image obtenue est représentée Figure II.12(b). On remarque qu'une partie du contour du poivron central a été supprimé. Les valeurs de seuils sont donc trop élevées pour cette image. Le choix de valeurs plus faibles serait mieux adapté comme l'illustre la Figure II.13(c). En effet, même si cette image contient quelques artefacts, pour chaque végétal, on obtient un contour connexe et complet. Nous en concluons qu'il n'existe pas de bonnes valeurs

de seuils, adaptées à toutes les images. Dans notre étude, nous avons constaté que les valeurs de seuils diffèrent d'une image à l'autre. Il est donc nécessaire de recalculer ces valeurs pour chaque image.

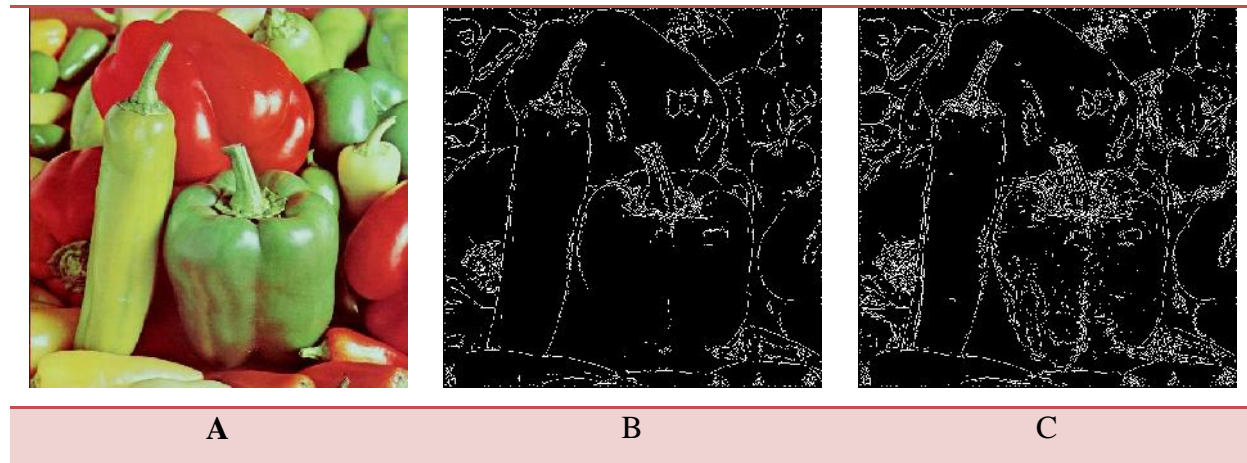


Figure II.13 : (a) Image originale des poivrons et détection de contours sur cette image avec (b) le même seuil que pour Léna $s_h = 29$ et $s_b = 26$ et (c) avec un seuil plus faible $s_h = 12$ et $s_b = 10$.

Un autre problème qui se pose, concerne le modèle de bruit présent dans les images. Habituellement, le bruit contenu dans les images est considéré comme uniforme. Ce bruit est indépendant de la position dans l'image et des valeurs des niveaux de gris des pixels. Cependant, dans certaines images, le bruit peut avoir un comportement non-uniforme. En imagerie médicale, par exemple, le bruit présent dans certaines images suit une distribution de Poisson. Lors de la détection de contours sur ces images, le seuillage par hystérésis échoue. Il est alors impossible de déterminer un seuil approprié à l'ensemble de l'image. [19]

Afin d'illustrer nos propos, nous avons généré une image de synthèse présentée Figure II.14(a). Cette image représente une rosace constituée de trois disques qui s'intersectent. Ces disques contiennent différents niveaux de gris. A cette image, nous avons ajouté un bruit de Poisson. Nous l'avons ensuite traité avec un filtre de deriche et nous avons cherché une valeur de seuil permettant de détecter complètement les contours de nos trois cercles. Notre objectif est atteint pour de faibles valeurs de seuils ($s_h = 10$ et $s_b = 8$). L'image (b) représente les contours obtenus pour ces valeurs de seuil. Cependant, si nos trois cercles sont correctement détectés, l'image contient également de nombreux contours non désirés. Ces

contours sont liés uniquement au bruit de Poisson. Si on augmente les valeurs de seuils (Figures I.14(c), (d) et (e)), ces artéfacts sont supprimés de l'image, mais les contours des disques le sont aussi. Le seuillage n'est donc pas efficace lorsque les images contiennent un bruit non-uniforme.

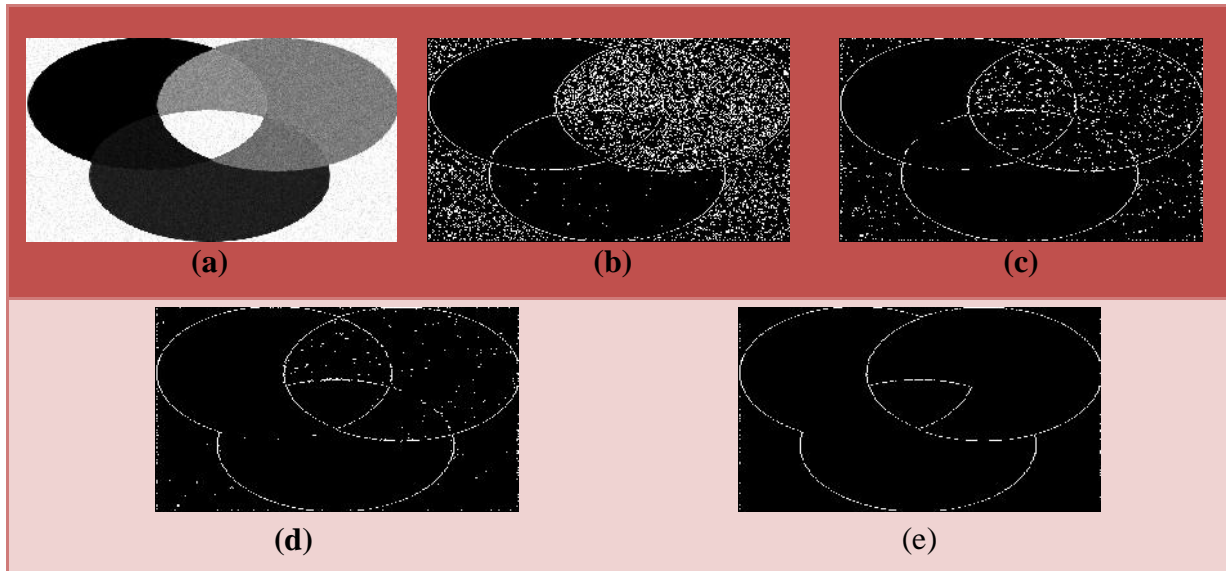


Figure II.14 (a) Image originale bruitée avec un bruit de Poisson. (b) $s_h = 10$ et $s_b = 8$,
(c) $s_h = 20$ et $s_b = 18$, (d) $s_h = 25$ et $s_b = 23$, (e) $s_h = 108$ et $s_b = 105$.

Nous avons pu observer à travers plusieurs exemples que les valeurs de seuils sont difficiles à définir. Elles diffèrent d'une image à l'autre et il n'y a pas d'approche universelle pour les fixer. De plus, il existe plusieurs sortes de bruits en fonction du capteur utilisé et pour les bruits non uniformes, les méthodes de seuillage classiques échouent. Pour résoudre ces problèmes, nous avons donc proposé une nouvelle approche d'estimation de gradient permettant de souiller l'image quelle que soit sa nature et le bruit qu'elle contient.

II.3.3.2 Segmentation de vidéo

Un algorithme de la segmentation découpe une vidéo de l'entrée dans plusieurs composants. Chaque composant est w.r.t homogène. Un ou plus de propriétés c.-à-d. la variation de dimensions dans les régions moins de variation devrait être aux frontières de l'objet considérablement. [20]

Les nombreux algorithmes de la segmentation existent dans la littérature. Aussi toute segmentation particulière l'approche ne peut pas être convenable pour

toute la vidéo tâches compréhensives. Approximativement 30 nouvelle segmentation les approches sont publiées à une conférence de la vision de l'ordinateur majeure, en soulignant le voûtez pour une approche de la segmentation fiable. Cependant une segmentation fiable et robuste encore élude la communauté de la vision.

Les exigences pour segmentation dépendent de la vidéo tâche compréhensive. De plus il peut y avoir des algorithmes de la segmentation différents qui satisfont la même compréhension de la vidéo la tâche. Un gens qui comptent le système peuvent utiliser un mouvement a basé segmentation ou une segmentation basé sur filtres dont la réponse peut localiser les personnes dans une vidéo.

Les approches de la segmentation de la vidéo peuvent être classées dans deux catégories largement, basé sur l'élan sur couleur sur les renseignements du mouvement obtenir une segmentation:

- Colore a basé: Les approches qui appartiennent à cette catégorie à l'origine comptent sur couleur des images élémentaires exécuter la segmentation.
- Motion a basé: Le mouvement fournit des indices vitaux pour grouper et algorithmes sous ceci la catégorie compte sur les renseignements du mouvement pour accomplir la segmentation principalement. À cause de la difficulté dans extraction de mouvement à couleur régions homogènes, beaucoup d'algorithmes, appartenir à cette catégorie peut fournir groupement de seulement un sous-ensemble d'images élémentaires dans une vidéo. La dernière ligne dans Chiffre 1.5 expositions un exemple pour mouvement clairsemé a basé le groupement. Ce groupement a peu ou aucune cohérence spatiale dans la production. Typiquement ce clairsemé les représentations fournissent le groupement pour un sous-ensemble simple de 3 à 5% des images élémentaires totales dans la vidéo.

La plupart de la segmentation approche utilisez une combinaison de mouvement et renseignements de la couleur. Cependant les segments obtenus sont 2D+time (3D) taches, a appelé comme supervoxels.

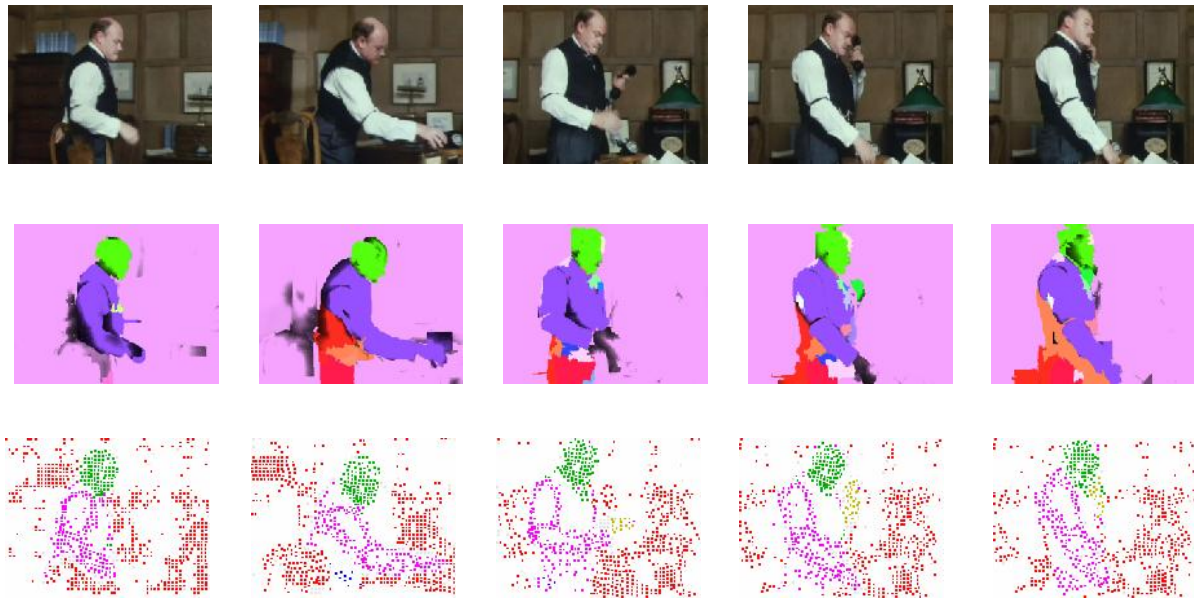


Figure II.15 : En premier Ligne: L'échantillon entré des cadres de l'image de marple13 classe par [15].

Ligne Centrale: Une segmentation de la vidéo dense. Les couleurs représentent différents groupes. La Ligne inférieure: Un groupement de trajectoires de l'image élémentaire (par [15]). Les zones blanches dans les images n'est pas étiquetée et d'où n'appartient pas à tous segments. [21]

Les quels fournissent peu ou aucun détail au sujet de spatial mouvement temporel au sujet d'images élémentaires qui constituent les tâches.

L'approche proposée dans cette thèse utilise couleur et renseignements du mouvement pour segmenter une vidéo. Par contraste avec la segmentation dernière existante approche, nous visons à construire long terme que le mouvement spatial - temporel détaille pour toutes les images élémentaires dans la vidéo.

La plupart des approches de la segmentation proposées jusqu'aux début 2000s est basé sur encadrer le traitement par - cadre. Récemment, avec l'augmentation dans efficacité computationnelle des ordinateurs, il y a une émergente tendance pour calculer des représentations basées en empilant de multiples cadres consécutifs (dans l'ordre de centaines) de la vidéo. Nous appelons ces volumes de la vidéo du tas. Les approches volume - basées ont gagné l'importance, comme traiter tous les cadres d'une vidéo conjointement plus de renseignements apportent et aides qui

maintiennent une segmentation cohérente avec le temps. **Figure II.16** tranches des expositions d'un volume de la vidéo formées en empilant 100 cadres consécutifs. Dans ce chiffre, la première ligne indique des cadres de l'échantillon d'une vidéo. L'image inférieure (encart) expositions le volume de la vidéo par lui est trois orthogonales découpes en tranche (cf. légende pour plus de détails).

Sur ce volume de la vidéo, une de la tâche de la segmentation la plus simple pourrait être extraire le premier plan en calculant des bords dans le temporel (T) dimension. L'intuition est que pour un fixe la vidéo de l'appareil-photo, si une personne marche à l'axe optique de l'appareil-photo avec suffisant perpendiculairement hâtez-vous, alors les images élémentaires qui correspondent aux bords appartiendront à la personne. À ceci terminiez nous utilisons les Deriche filtre [6] pour calculer des bords le long de l'axe temporel. Ce filtrage est suivi par un pas du thresholding où bords suffisamment fort estimés est resté et le reste est supprimé (zones sombres dans la figure II.17). Donc nous obtenons maintenant un binaire étiqueter de la vidéo où le premier plan correspond aux zones blanches dans la figure II.17. La haute cohérence dans origine étiqueter est dûe à traiter le 2D+T données conjointement dans la forme d'un volume de la vidéo. Ces aides du traitement communes un détecteur du bord localiser



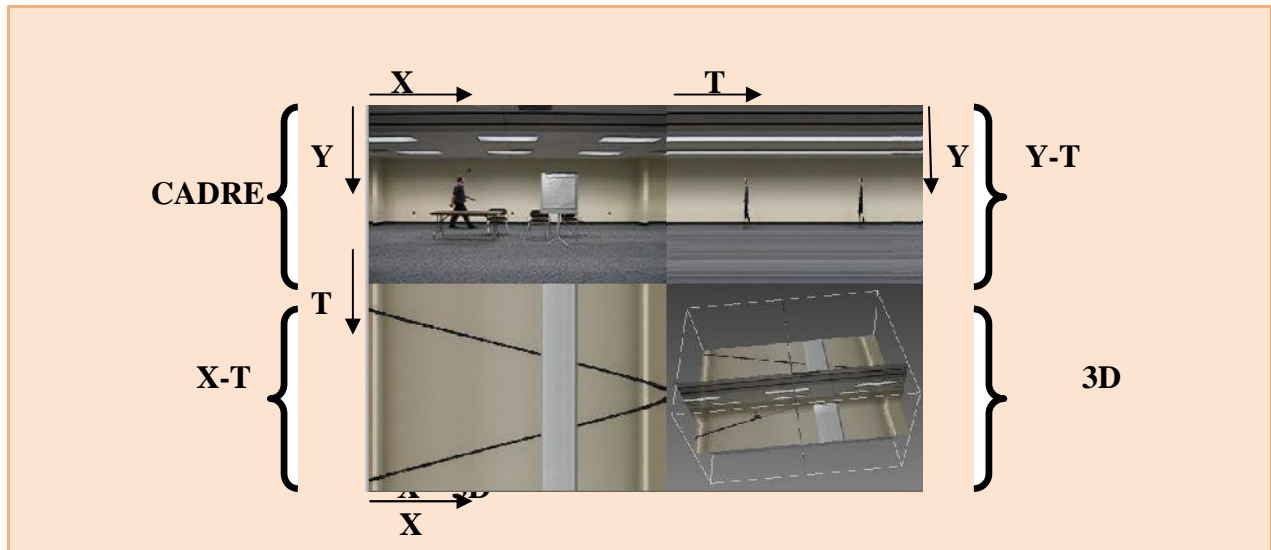


Figure II.16 : Tranches d'un volume de la vidéo. La ligne du sommet montre quelques cadres de l'échantillon d'une vidéo.

Expositions de la ligne inférieures que les trois orthogonales découpe en tranche du volume. La ligne bleue indique la séparation des tranches. Les expositions de l'encart du droit inférieures l'emplacement de trois tranches à l'intérieur de volume. Bords dans un meilleur chemin que considérer seulement deux cadres consécutifs.

Un autre avantage apporté par ce volume est des renseignements de la gamme temporels longtemps. Local les renseignements du mouvement parmi spatialement images élémentaires avoisinantes peuvent être inexacts comme montré dans Chiffre 1.3. La cependant modelant longue affinité du terme (avec les critères convenables) entre images élémentaires pourrait aider dans arriver ce problème à couleur régions homogènes autour. Avec référence à l'exemple dans Chiffre 1.3, une approche viable construire le courant exact dans les zones rouges (manquer de structure de l'intensité discriminante) est propagé le haut courant de l'exactitude des bords et coins sur la personne en mouvement. Dans cette thèse nous poursuivons des volumes de la vidéo pour obtenir une segmentation de la vidéo robuste, fiable et dense avec longue évaluation du mouvement de la gamme pour chaque image élémentaire dans la vidéo.

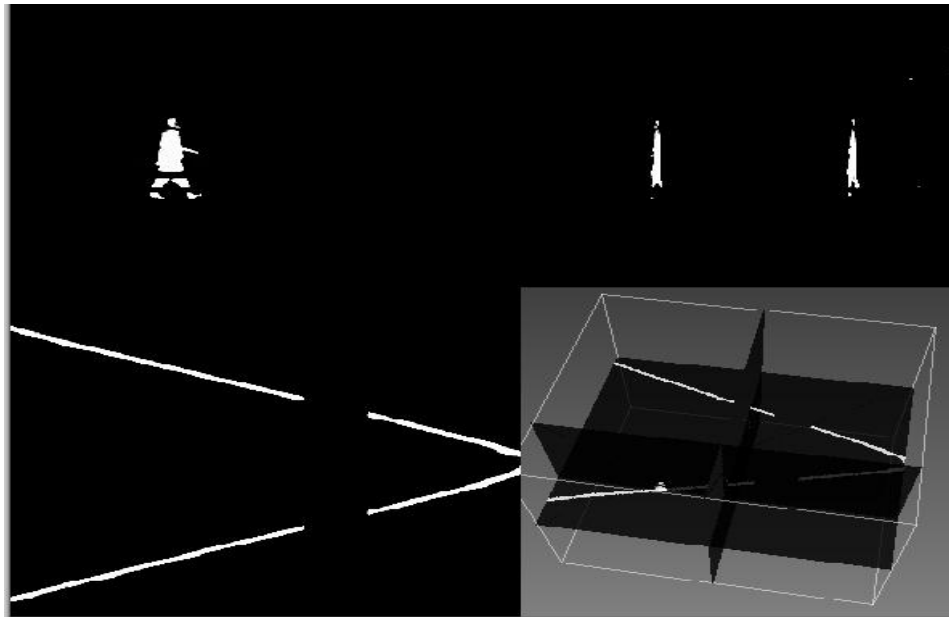


Figure II.17 : le filtre Deriche est appliqué sur un volume de la vidéo pour extraire le premier plan (Blanc les images élémentaires constituent le premier plan).

II.4 Conclusion

La segmentation est une étape primordiale en traitement d'image et du fait qu'il n'existe pas de méthode unique de segmentation, il est difficile de définir, de manière absolue, une « bonne » segmentation. La segmentation n'est pas une fin en soi, sa qualité est fonction des résultats obtenus par les traitements situés en aval qui utilisent les primitives extraites. [22]

Ce chapitre dresse brièvement un état d'art des principales méthodes de segmentation. Chacune possède ses avantages et ses inconvénients : le problème majeur de la segmentation basée sur les contours est la fermeture des contours, d'autre part, la segmentation basée sur les régions peut s'avérer « gourmande » en calcul. En l'absence de contextes particuliers,

Puisque la qualité du résultat est principalement Subjective et que la finalité d'utilisation de la segmentation obtenue influe sur la perception de qualité de ce résultat, le choix d'une technique plutôt qu'une autre est assez difficile.

Chapitre III

Processus de détection de texte sur les séquences vidéo et les critères dévaluations

III.1 Introduction

Depuis quelques années, les documents audiovisuels numérisés sont de plus en plus fréquents. Des grandes bases de données audiovisuelles ont été créées par des entreprises, des organisations et aussi par des personnes privées. Cependant, l'utilisation de ces bases reste encore problématique. Tout particulièrement, ces nouveaux types de données, image et vidéo, ont conduit à de nouveaux systèmes d'indexation ou la recherche par le contenu se fait à partir d'une image exemple. Les systèmes disponibles actuellement travaillent sans connaissance (systèmes pré-attentifs).

Ils utilisent des méthodes de traitement d'images pour extraire des caractéristiques de bas niveau (couleur, texture, forme etc.). Malheureusement les requêtes construites à partir de ces caractéristiques ne correspondent pas toujours aux résultats obtenus par un humain qui interprète le contenu du document. Cette étude est soutenue par France Telecom Recherche et Développement dans le cadre du projet ECAV. La cause de cet échec est le manque de sémantique dans l'information extraite.

Considérons l'image de la figure 1a. Un utilisateur humain qui choisit cette image comme image requête est probablement intéressé par des images de cyclistes (notons, que nous ne pouvons pas deviner le vrai désir de l'utilisateur et par conséquent, la similarité entre deux images n'est donc pas accessible). D'autre part les systèmes d'indexation trouvent des images qui sont similaires par rapport aux caractéristiques de bas niveau.

Les images contenant des cyclistes qui ne sont pas similaires par rapport à ces caractéristiques (voir figure 1b), ne seront pas trouvées.

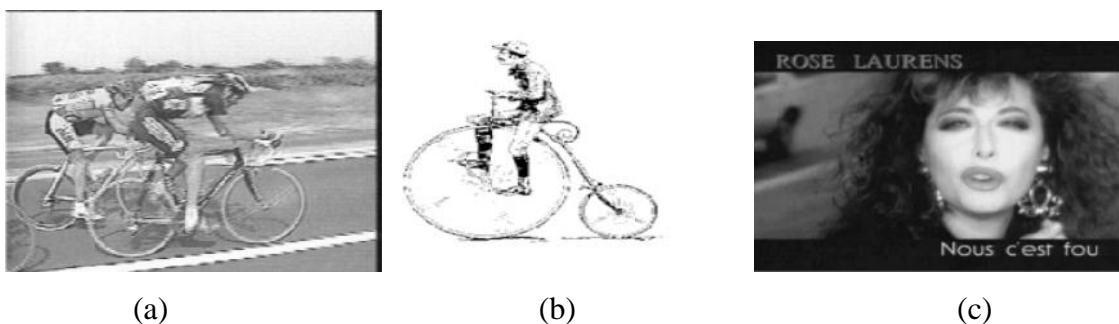


Figure III.1 : Images d'exemples.

Entre le niveau du pixel, et celui de la sémantique, il existe des caractéristiques à la fois riche en information et cependant simples. Le texte présent dans les images et les vidéos fait

partie de cette catégorie. La Figure 1c montre une image extraite d'une publicité fournie en requête par un utilisateur. On peut imaginer que celui-ci cherche l'image de la chanteuse.

La sémantique de cette image exemple peut être résumée par : Une publicité pour le nouvel album de la chanteuse Rose Laurens avec une photo de son visage". Il est impossible de déduire cette information à partir des caractéristiques basses. Pourtant dans ce cas le nom de la chanteuse est présent dans l'image. Par conséquent, en extrayant le texte on obtient une information supplémentaire très valable. Les requêtes classiques peuvent donc être complétées par des mots clés. Le texte automatiquement extrait peut être stocké. Le processus que nous étudions dans ce travail concourt donc à une meilleure automatisation de la phase d'indexation d'images et de vidéos.

III.2 Formulation du problème

Les recherches en détection et l'extraction du texte à partir des séquences vidéo sont encore confrontées à des sérieux problèmes. Le problème principal peut être expliqué par la différence entre l'information présente dans un document et celle donnée par une séquence vidéo ainsi que les méthodes de stockage de chaque type. Nous allons décrire ces divergences dans les paragraphes suivants. Pour améliorer la qualité et le taux de la reconnaissance, les images sont numérisées à très haute résolution (200-400 dpi) donnant des fichiers de taille très élevée (un fichier de 100 Mo résultant d'une page A4 numérisée à 400dpi). Les fichiers sont comprimés sans perte pour garder la qualité et empêcher des artefacts de compression. Ces grandes tailles ne sont pas une limite puisque ni leur transport ni leur stockage ne sont prévus. La plupart du temps les images de documents sont bien structurées et contiennent un fond uniforme et la couleur du texte est également uniforme. Ceci permet de séparer les caractères du fond avec un seuillage fixe ou adaptatif.

La majorité de la page contient des caractères structurés dans différents paragraphes, quelques images sont incluses dans la page. Par contre les images des séquences vidéo contiennent de l'information plus difficile à traiter. Le texte n'est pas séparé du fond. Il est soit superposé (le texte artificiel" comme les sous-titres, les résultats du sport etc.) soit inclus dans la scène de l'image (le texte de scène", par exemple le texte sur le tee-shirt d'un acteur). Le fond de l'image peut être très complexe ce qui empêche une séparation facile des caractères. De plus, contrairement aux documents écrits, les séquences vidéo contiennent de l'information très riche en couleurs.

Enfin, le texte n'est pas structuré en lignes, et souvent quelques mots courts et déconnectés sont présents dans l'image.

Le but principal de la conception des formats de fichiers vidéo est de garder une qualité suffisante pour l'archivage, pour le stockage et le transport par des réseaux informatiques.

Pour cette raison et pour une quantité de données vidéo plus hautes, il est nécessaire de réduire fortement l'information avant son stockage dans le fichier vidéo.

Pour limiter la taille de l'information deux méthodes sont souvent appliquées : la forte réduction de la résolution et le codage avec perte, c.à.d. avec élimination des données redondantes et perte d'une partie de l'information originale, en utilisant les algorithmes de compression JPEG et MPEG.

III.3 Etat de l'art

L'extraction du texte des flux de vidéo est un domaine de recherche très récent mais cependant très fécond. La plupart des travaux traitent le problème de la détection et de la localisation du texte (nommée "détection") ainsi, il existe très peu de recherche sur la reconnaissance, surtout à cause des problèmes présentés dans la section 2.

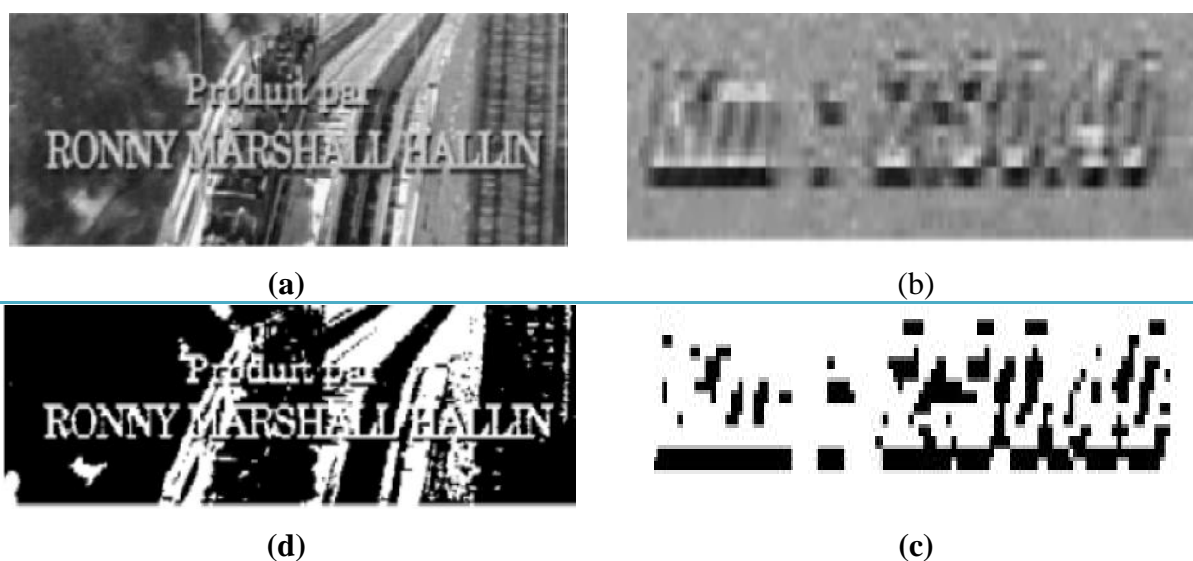


Figure : III.2 : Texte sur un fond complexe (a) un essai de seuillage fixe (b) Rehaussement artificiel du contraste (c) un essai de seuillage fixe (d)

La problématique de la détection est en émet proche de celle de l'analyse de la structure d'un document contenant du graphisme (Journaux, Page web etc.).

A titre l'exemple, on peut citer les travaux de Jain et al.

En supposant que la couleur des caractères est uniforme, ils utilisent une réduction des couleurs suivie par une segmentation et une phase de regroupement spatial pour trouver les caractères du texte. Bien que la présence des caractères joints est prévue par les auteurs, la phase de segmentation pose des problèmes dans le cas des documents de mauvaise qualité et notamment les séquences vidéos de basse résolution. [23]

Une approche similaire, qui donne des résultats impressionnants sur le texte de grande police, est présentée par Lienhart et al. D'ailleurs ils sont parmi les premiers qui ont présentée une étude aussi complétée partant de la détection jusqu'à la reconnaissance et l'indexation.

Étant mieux adaptée aux propriétés de la vidéo, l'algorithme proposé combine une phase de détection par segmentation des caractères et la détection par une recherche de contraste local élevée. Ensuite une analyse de texture supprime les fausses alarmes. Le suivi du texte est effectuée au niveau des caractères. Pour la reconnaissance les auteurs utilisent un produit commercialisme. Malheureusement ils ne pouvaient pas montrer l'aptitude de l'algorithme de segmentation appliquée aux textes de petite taille. [24]

La méthode de Sato, Kanade et al. Repose sur le fait que le texte est composé de traits de contraste élevée. L'algorithme de détection cherche des contours groupés en rectangles horizontaux. [25]

La reconnaissance est réalisée par une mesure de corrélation. Wu, Manmatha et Riseman combinent la recherche des contours verticaux et l'application d'un filtre de texture. Le travail de Le Bourgeois est basé sur l'accumulation des gradients horizontaux. La reconnaissance dans le système proposée est exécutée par des règles statistiques sur les caractéristiques de projections des niveaux de gris. Li et Doberman présentent une approche d'apprentissage En laissant glisser une sous-image sur l'image, ils utilisent un réseau de neurones de type MLP pour classer chaque pixel comme texte ou non-texte. Les caractéristiques sont extraites des échelles de hautes fréquences d'une ondelette de Haar. Clark et Mirmehdi utilisent aussi un réseau de neurones pour la détection du texte. Ils extraient des caractéristiques diverses comme la variance de l'histogramme, la densité des contours etc. [26]

Une méthode qui travaille directement dans le domaine de la vidéo comprimée est proposée par Zhang, Zhang et Jain. Leurs caractéristiques sont calculées directement à partir des coefficients de DCT des données MPEG. [27]

La plupart des problèmes est identifiée, mais pas encore résolue. Surtout, l'aspect d'amélioration du contenu, bien qu'aborde par Sato, Kanade et al. [3] et par Li et Doermann [6], ne donne pas encore des résultats tout a fait satisfaisants. Nous considérons aussi la segmentation des caractères du fond complexe comme un problème de grande importance, qui demande encore une recherche intensive. [28]

III.4 Un système d'extraction

Nous pouvons décrire le but principal d'un système d'extraction de texte par quelques mots : accepter des fichiers d'images et de vidéo, détecter le texte, l'extraire et produire un fichier ASCII incluant le texte dans un format utilisable pour le traitement. En tenant compte de la problématique expliquée précédemment, la structure d'un système proposée dans la figure 3 s'impose naturellement.

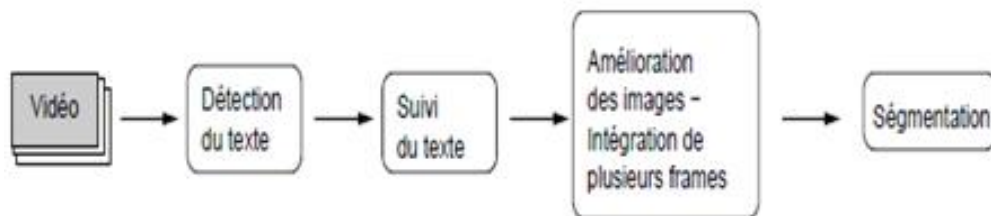


Figure III.3 : Le fonctionnement général du système. [29]

La détection du texte est réalisée dans chaque frame de la vidéo.

1. Les rectangles figurant la localisation du texte sont suivis pendant leur période d'apparition pour associer les rectangles se correspondant dans les différents frames. Cette information est nécessaire pour améliorer le contenu de l'image, ce qui peut être atteint par l'intégration de plusieurs rectangles contenant le même texte... Par conséquent il est aussi nécessaire d'augmenter la résolution, en utilisant l'information supplémentaire prise dans la séquence d'images. Dans cette section nous proposons un système de détection et de suivi du texte. Notre système est capable de détecter et localiser le texte dans une image ou dans un frame, et de suivre les occurrences du texte dans une séquence vidéo. Nous ne présenterons que peu de détails dans cette section (uniquement les principes généraux) car cette recherche est effectuée dans le cadre d'un contrat et les premiers résultats donnent lieu à dépôt de brevets.

2. Les perspectives que nous pouvons envisager pour notre projet sont donc fondées sur le développement d'un système complet qui sera capable d'extraire le texte, de le reconnaître et enfin de l'utiliser pour l'indexation. [30]

III.5 Les Critères d'évaluations

Devant le foisonnement de méthodes développées depuis plusieurs décennies pour la segmentation des images, le problème de l'évaluation est devenu crucial.

Disposer de méthodes d'évaluations de résultats est nécessaire :

- Aux chercheurs pour comparer un nouvel algorithme à ceux préexistants
- Aux utilisateurs pour choisir un algorithme et régler ses paramètres en fonction du problème à résoudre.

Les critères d'évaluation quantitative peuvent être groupés en deux classes, selon que l'on possède ou non une "vérité-terrain" qui constitue une segmentation de référence. Celle-ci est directement accessible dans le cas d'images de synthèse, mais elle doit être construite "à la main" par un expert du domaine de l'application dans le cas d'images réelles : traces effectuées par des médecins, des géographes, etc. à l'aide d'outils informatiques de dessin.

Si l'on veut comparer de manière objective les méthodes, il est plus simple d'utiliser des images de synthèse, pour lesquelles une "vérité" est parfaitement connue, à savoir la segmentation qui a servi à synthétiser l'image. L'inconvénient d'une telle démarche est que ces images ne représentent pas toutes les situations possibles d'une utilisation réelle.

Bien que l'évaluation sur images réelles soit certainement plus réaliste, elle pose d'autres difficultés, la principale étant qu'il n'existe généralement pas de solution unique à la division d'une image en régions "pertinentes".

La "pertinence" d'une région est en effet une notion éminemment dépendante de l'application : qui cherche à suivre les individus dans une vidéo sera intéressé par le détournement des silhouettes, qui s'intéresse à la mode voudra isoler les vêtements des personnages, etc. Néanmoins, deux segmentations humaines d'une même image tendent à être cohérentes dans le sens où elles sont des raffinements mutuels l'une de l'autre [31]: certaines régions d'une segmentation constituent une sur-segmentation de certaines régions de l'autre et inversement.

Autrement dit, la principale différence entre deux segmentations humaines d'une même image est une différence de niveau de détail. Yu et Shi [32] ont récemment proposé une classification des méthodes de segmentation en deux grandes catégories : d'une part une approche dite "discriminative" qui, dans la droite ligne des méthodes de classification non

supervisée, envisage la segmentation comme un problème de regroupement des pixels en classes compactes et bien séparées et d'autre part une approche dite "générative", de type "problème inverse", qui envisage la segmentation comme un problème de recherche d'un "modèle générateur" des données (générative model). Or le problème de l'évaluation de la segmentation sans référence est extrêmement proche du problème de segmentation lui-même. Nous verrons que les critères d'évaluation sans référence suivent cette même classification en deux grands groupes, l'un basé sur l'approche "discriminative" aboutit à des mesures de type contraste et l'autre basé sur l'approche générative, conduit à une modélisation par morceaux de l'image.

Quand on dispose d'une vérité-terrain, l'évaluation des segmentations s'effectue à l'aide de critères comparant chaque segmentation avec l'image de référence. On peut ainsi ordonner les segmentations.

En l'absence de vérité-terrain, il faudra employer des critères quantitatifs absolus ou des calculs de cohérence entre les différents résultats de segmentation.

Dans la suite de l'article, une image I est définie sur un ensemble de sites X représentant les coordonnées spatiales des pixels (ligne, colonne) et une fonction f à valeurs dans un ensemble Z . Par exemple f pourra être l'intensité pour les images à niveaux de gris (dans ce cas Z est un sous-ensemble de N) et la couleur dans l'un des espaces colorimétriques pour les images couleurs (dans ce cas Z est un sous-ensemble de N^3).

On notera R une segmentation à évaluer, c'est donc une partition de X en régions notées R_i , $i = 1, \dots, N$ vérifiant

$$R_i \cap R_j = \emptyset \text{ et } \bigcup_{i=1}^N R_i = X \quad (1)$$

On note A le nombre de pixels de l'image et A_i le nombre de pixels de la région R_i . On a donc :

$$A = \sum_{i=1}^N A_i = \text{card } X \quad (2)$$

III.5.1 Avec segmentation de référence

Dans cette partie, nous disposons d'une segmentation de référence notée V , dont les régions sont notées V_i , $i = 1, \dots, M$.

III.5.1.1 Mesure de Vinet

La mesure de Vinet s'appuie sur un appariement biunivoque entre les régions des deux segmentations à comparer. Pour tout couple de régions (V_i, R_j) , on définit leur recouvrement par $t_{ij} = \text{card}(V_i \cap R_j)$. Un couplage de poids maximal du graphe bipartite (V, R, t) fournit alors un appariement biunivoque optimal entre régions des deux segmentations au sens de la somme des recouvrements des parties appariées. Soit K le nombre de couples obtenus et $C_1 \dots C_K$ les recouvrements de chacun de ces couples.

$$\frac{1}{A} \times \sum_{k=1}^K C_k \quad (3)$$

C_k représente alors le poids total du couplage normalisé par la surface de l'image. La mesure de dissimilarité de Vinet est alors.

$$1 - \frac{1}{A} \times \sum_{k=1}^K C_k \quad (4)$$

Bien que le couplage ne soit pas nécessairement unique cette mesure est une distance. Une approximation de la mesure de Vinet peut être obtenue efficacement par un algorithme glouton qui consiste à coupler itérativement les deux régions de recouvrement maximal. Cette mesure de dissimilarité a été utilisée dans [7] pour comparer des segmentations sur des images synthétiques monochromes comportant différents bruits et textures.

Des mesures généralisant la distance de Vinet à des appariements multivoques ont été proposées dans [4] et [9].

III.5.1.2 Mesure de cohérence entre segmentations de Martin. [33]

Définie par D. R. Martin pour évaluer la cohérence entre deux segmentations manuelles d'une même image, cette mesure peut être utilisée pour comparer deux segmentations l'une de référence, l'autre obtenue par un algorithme.

Elle est basée sur deux erreurs calculées en chaque pixel : une erreur de V par rapport à R et une erreur de R par rapport à V . Si le pixel s appartient à la région V_j dans la vérité-terrain et à la région R_i dans l'image résultat, ces erreurs valent :

$$E(c) = \frac{\text{card}(V_j \setminus R_i)}{\text{card}(V_j)} \text{ et } E'(s) = \frac{\text{card}(R_i \setminus V_j)}{\text{card}(R_i)} \quad (5)$$

$E(s)$ vaut 0 si V_j est un sous-ensemble de R_i et vaut 1 si l'intersection des deux régions réduite au pixel s . La dissimilarité entre segmentation résultat et segmentation de référence se mesure alors par l'erreur locale de cohérence :

$$LCE(R, V) = \frac{1}{A} \sum_s \min\{E(s), E'(s)\} \quad (6)$$

Ou par l'erreur globale de cohérence :

$$GCE(I, V) = \frac{2}{A} \min \{ \sum_s E(s), \sum_s E'(s) \} \quad (7)$$

Cette dernière mesure est plus sévère que la première et a le défaut de favoriser une segmentation (ou une sous-segmentation) de toute l'image par rapport à un mélange des deux situations (sur et sous-segmentation selon les zones de l'image).

III.5.1.3 Position des pixels mal segmentés : mesure de Yasnoff et al. [34]

Compter simplement le nombre de pixels mal segmentés est insuffisant, il faut aussi tenir compte de la position de ces pixels en utilisant par exemple la distance entre un pixel mal segmenté et la région à laquelle il appartient dans la référence.

La mesure de Yasnoff et al. s'écrit :
$$\frac{100}{A} \times \sqrt{\sum_s d^2(s)} \quad (8)$$

Où la sommation s'effectue sur les pixels mal segmentés et d est la distance au pixel le plus proche de la région à laquelle il appartient. Cet indice a été utilisé dans [35] pour comparer des méthodes de seuillage.

III.5.2 Sans segmentation de référence

De nombreux critères ont été proposés, cherchant à quantifier la qualité ou la lisibilité de l'image. Suivant la classification des méthodes de segmentation proposée dans [36], on peut classer ces critères en deux grandes catégories : les critères de "contraste" et les critères d'adéquation à un modèle". Les premiers recherchent une variabilité inter-région, alors que les seconds recherchent une uniformité en intensité ou en couleur à l'intérieur des régions.

Parmi les critères de contraste, nous présentons ci-dessous celui de Levine et Nazif et celui de Zeboudj, ainsi que le critère de Rosenberger. Parmi les seconds, nous présentons le critère d'uniformité de Levine et Nazif, le critère de Liu et Yang, et celui de Borsotti et al.

Si f représente un attribut du pixel (en général son intensité ou sa couleur), on notera m_i (resp. σ_i) La moyenne (resp. l'écart-type) de f dans la région R_i .

III.5.2.1 Contraste inter-région de Levine et Nazif [36]

Soit $C_{ij} = \frac{|m_i - m_j|}{m_i + m_j}$ le contraste entre deux rag i, j sur le périmètre de R_i . (9)

Le contraste global est alors :

$$\frac{\sum_{R_i} w_i C_i}{\sum_{R_i} w_i} \quad (10)$$

w_i est un poids associé à chaque région, qui peut être l'aire de la région. Ce critère, utilisé par Zhang [13] sans pondération s'est révélé absolument non discriminant.

III.5.2.2 Contraste de Zeboudj

Cet indice prend en compte le contraste intérieur et le contraste extérieur aux régions,

Mesures sur un voisinage $W(s)$ du pixel s . $C(s, t) = \frac{|f(s) - f(t)|}{L - 1}$ (11)

Soit le contraste entre deux pixels s et t , avec f représentant l'intensité et L le maximum des intensités.

Le contraste intérieur d'une région R_i est :

$$I_i = \frac{1}{A_i} \sum_{s \in R_i} \max \{c(s, t), t \in W(s) \cap R_i\} \quad (12)$$

Le contraste extérieur d'une région R_i est :

$$E_i = \frac{1}{l_i} \sum_{s \in F_i} \max \{c(s, t), t \in W(s), t \notin R_i\} \quad (13)$$

Où F_i est la frontière de R_i et l_i la longueur de F_i .

Le contraste de R_i est :

$$C(Ri) = \begin{cases} 1 - \frac{Ii}{Ei} & \text{si } 0 < Ii < Ei \\ Ei & \text{si } Ii = 0 \\ 0 & \text{sinon} \end{cases} \quad (14)$$

Le contraste global est enfin :

$$\frac{1}{A} \sum_i Ai \cdot C(Ei) \quad (15)$$

Pour résoudre le problème des images monochromes contenant des textures, Rosenberger commence par caractériser chaque région en région texturée ou uniforme, Grâce à un calcul d'uniformité des niveaux de gris basé sur les matrices de cooccurrences. Il calcule ensuite la disparité intra-région, notée D et la disparité inter-région, notée D . La première correspond à l'écart-type des intensités pour une région uniforme et a un ensemble d'attributs de texture pour une région texturée. D est égale à la différence des moyennes pour deux régions uniformes, à la distance euclidienne entre attributs de texture pour deux régions texturées et à 1 pour une région texturée et une région uniforme. La disparité intra-région globale est égale à la moyenne pondérée des disparités calculées pour chaque région :

$$D = \frac{1}{N} \sum_i Ai \cdot C(Ei) Di \quad (16)$$

Et de même pour la disparité inter-région globale. Finalement le critère de Rosenberger est égal à :

$$\frac{\bar{D} - D}{2} \quad (17)$$

III.6 Les operateurs de segmentation

La détection de contours est très utilisée comme étape de prétraitement pour la détection d'objets, pour trouver les limites de régions. En effet, un objet peut être localisé à partir de l'ensemble des pixels de son contour. De plus, trouver cet ensemble permet d'obtenir une information sur la forme de l'objet. Du point de vue théorique, un contour est défini par un changement marqué de l'intensité d'un pixel à l'autre. Comme en chaque point, le contour est considéré comme perpendiculaire à la direction du gradient, ce dernier est utilisé pour la

détection : selon les modèles classiques de traitement d'image, en tout point de l'image les dérivées partielles sont estimées. Ceux qui fournissent les réponses les plus fortes correspondent à des pixels de contour.

En pratique, le calcul de gradient est effectué par la convolution de filtres linéaires. Ces filtres peuvent estimer les dérivées premières, les dérivées secondes, etc., et se focaliser sur des contours de directions différentes. Il existe ainsi de nombreux filtres, qui ont des fondements mathématiques différents. Cependant, quelle que soit leur justification théorique, leur objectif reste le même : mettre en évidence les pixels qui ont une valeur très différente de leurs voisins. Nous ne présentons ici, à titre d'exemple, que le filtre gradient et les opérateurs de Robert et Prewitt parmi les plus couramment employés.

III.6.1 Technique du Gradient

Soit la fonction image $F(x, y)$, le gradient de F en un point (x, y) est un vecteur bidimensionnel qui s'écrit :

$$\vec{G} = \begin{bmatrix} \frac{\partial F}{\partial X} \\ \frac{\partial F}{\partial Y} \end{bmatrix} \quad (18)$$

Le gradient \vec{G} est calculé en chaque point (x, y) de l'image, son orientation va dans le sens de la variation maximale de F et sa norme est généralement donnée par :

$$|\vec{G}| = \left(\left(\frac{\partial F}{\partial X} \right)^2 + \left(\frac{\partial F}{\partial Y} \right)^2 \right)^{1/2} \quad (19)$$

Localement, la norme du gradient est d'autant plus élevée que les courbes de niveau sont rapprochées et de valeurs différentes. La direction de G par rapport à l'axe des X est

donnée par l'angle :

$$\theta = \arctg \left(\frac{\partial F / \partial Y}{\partial F / \partial X} \right) \quad (20)$$

Le calcul du gradient est utile pour mettre en évidence les contours. Le point (x, y) est un contour si la norme est supérieure à un seuil.

III.6.2 Opérateur de Robert

Le Roberts l'opérateur en colère fournit une simple approximation à la magnitude de l'inclinaison:

$$G[f[i,j]] = |f[i,j] - f[i+1,j+1]| + |f[i+1,j] - f[i,j+1]| \quad (21)$$

Utiliser la circonvolution masque, cela devient :

$$G[f[i,j]] = |G_x| + |G_y| \quad (22)$$

Où G_x et G_y sont calculés utiliser les masques suivants:

Tableau III.1 : Masques utilisés par Roberts Operateur.

$G_x =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px 15px;">1</td><td style="padding: 5px 15px;">0</td></tr> <tr><td style="padding: 5px 15px;">0</td><td style="padding: 5px 15px;">-1</td></tr> </table>	1	0	0	-1	$G_y =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px 15px;">0</td><td style="padding: 5px 15px;">-1</td></tr> <tr><td style="padding: 5px 15px;">1</td><td style="padding: 5px 15px;">0</td></tr> </table>	0	-1	1	0
1	0										
0	-1										
0	-1										
1	0										

Comme avec les 2 x antérieurs 2 opérateur de l'inclinaison, les différences sont calculées au point interpolé $[i + 1/2, j + 1/2]$. Le Roberts l'opérateur est une approximation à l'inclinaison continue au point interpolé et pas au point $[i, j]$ comme il peut être attendu. Comme par le Roberts Filtres de la Découverte du Bord, l'image du Requin, est montré dans le Figure III.4.

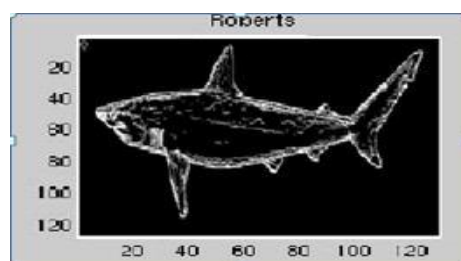


Figure III.4 : Image détectée par filtre de Roberts

III.6.3 Opérateur de Sobel.

Une façon d'éviter d'avoir l'inclinaison calculée au sujet d'un point interpolé entre les images élémentaires qui sont utilisées 3 x 3 voisinages pour les calculs de l'inclinaison [37]. Sur Le l'arrangement d'images élémentaires est au sujet de l'image élémentaire [i, j] montré dans la Tableau II.2. L'opérateur Sobel est la magnitude de l'inclinaison calculée par:

$$M\sqrt{S_x^2 + S_y^2} \quad (23)$$

Où les dérivés partiels sont calculés par:

$$S_x = (\alpha_2 + c\alpha_3 + \alpha_4) - (\alpha_0 + c\alpha_1 + \alpha_6) \quad (24)$$

$$S_y = (\alpha_0 + c\alpha_1 + \alpha_2) - (\alpha_6 + c\alpha_5 + \alpha_4) \quad (25)$$

Comme les autres opérateurs de l'inclinaison, S_x et S_y peuvent être rendus effectif utiliser des masques de la circonvolution:

Tableau III.2 : Masques utilisés par Opérateur de Sobel.

$S_x =$	1	0	1	$S_y =$	1	2	1
	-2	0	2		0	0	0
	1	0	1		-1	-2	-1

Notez que cet opérateur est placé sur une image élémentaires de l'emphasing c'est plus proche au centre du masque. Le Sobel l'opérateur est un du bord le plus communément utilisé les détecteurs.

Tableau III.3 : L'étiquetage d'images élémentaires de voisinage.

a_0	a_0	a_2
a_6	[i, j]	a_3
a_7	a_5	a_4

Comme par les Sobel Bord Découverte Filtrés, l'image du Requin est montrée dans la figure suivante :

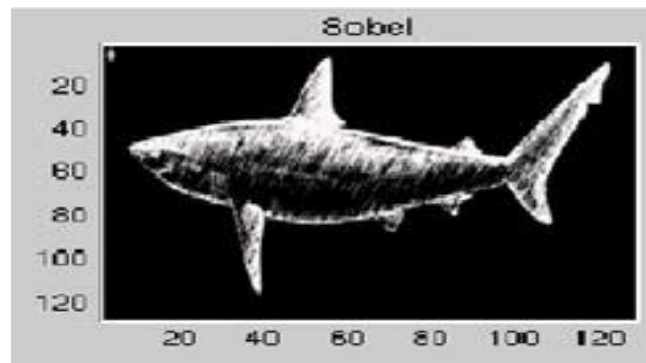


Figure III.5 Sobel Bord Découverte Filtrés.

La prochaine paire d'images est montrée dans l'horizontal et les bords verticaux ont sélectionné hors des images du requin du groupe avec la méthode Sobel de découverte du bord. Maintenant vous voulez observer la difficulté qu'il avait avec les certains traits du requin, tel que les branchies, bouche, nageoires et queues de requins différents comme montré dans le Figure III.6.

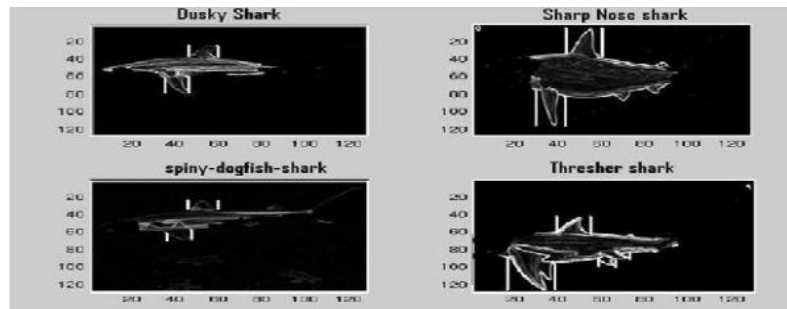


Figure III.6 : Filtre Sobel Vertical.

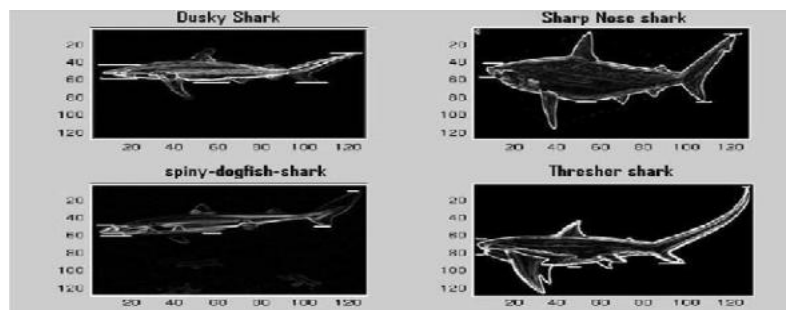


Figure III.7 : Filtre Sobel Horizontal.

III.6.4 Operateur de Prewitt

L'opérateur Prewitt utilise les mêmes équations comme le L'opérateur Sobel où c constant =

Tableau III.4 : Masques utilisés par Prewitt inclinaison Opérateur

$S_x =$	-1	0	1	$S_y =$	1	1	1
	-1	0	1		0	0	0
	-1	0	1		-1	-1	-1

Par conséquent, notez qui, contrairement à l'opérateur Sobel, ce l'opérateur ne place pas toute accentuation sur images élémentaires qui sont plus proche au centre des masques. [38]

Comme par le Prewitt la Découverte du bord Filtre, l'image du Requin est montrée dans le Figure suivant :

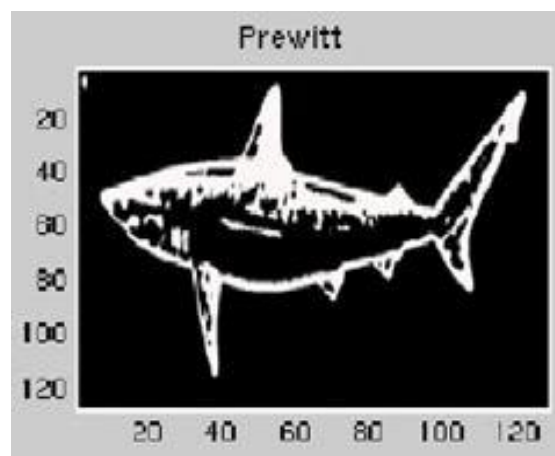


Figure III.8: Prewitt edge detection filter.

III.6.5 Le Laplacien

Le Laplacien d'une fonction $A(x, y)$ est défini de la manière suivante :

$$\Delta A(x; y) = \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} \quad (26)$$

Le Laplacien est souvent utilisé en amélioration d'images pour accentuer l'effet de contour. Il est caractérisé par une sensibilité au bruit accrue par rapport au gradient.

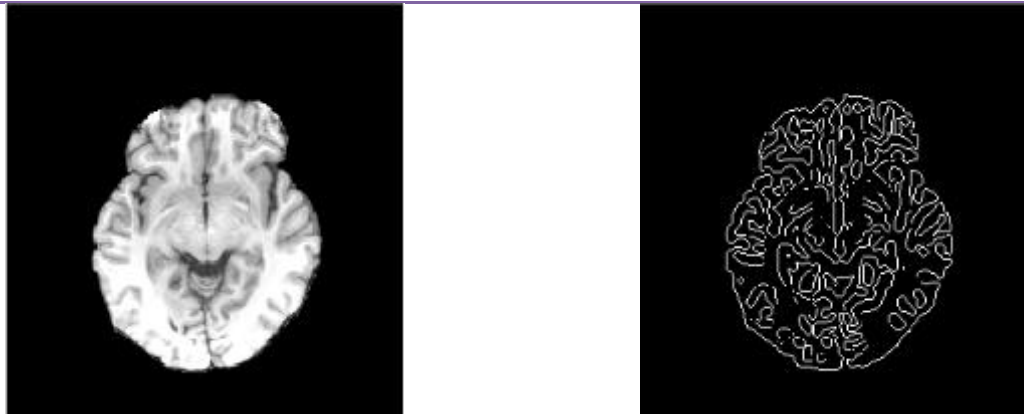


Image original

Détection de contours Laplace.

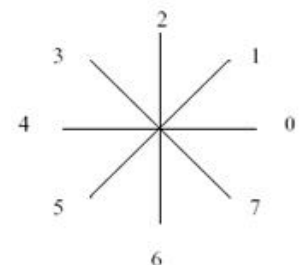
Figure III.9 : Détection de contours par Laplace.

III.6.6 Opérateur de Kirsh

Kirsh a proposé une méthode fondée sur l'utilisation de 8 masques.

L'opérateur de base h_0 est donnée par :

$$h_0 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & 3 & -3 \end{bmatrix} ; \quad h_0 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & 3 & -3 \end{bmatrix}$$



Codage de Freeman

Les autres opérateurs sont obtenus par rotation de $\pi/4$. [39]

III.6.7 Critère de Canny

Pour évaluer et comparer les performances relatives des divers détecteurs de contours existant, Canny a publié en 1986 trois critères numériques à optimiser pour obtenir la réponse impulsionnelle d'un filtre linéaire monodimensionnel qui soit le meilleur détecteur de contours possibles, selon ces critères, et qu'il adapte ensuite en

2 dimensions. Canny établit les bases de travail et définit un cadre dans lequel ses critères prennent un sens. Canny a donc proposé dans une approche originale, un filtre déterminé analytiquement à partir de trois critères :

1- Garantir une bonne détection, c'est-à-dire une réponse forte même à de faibles contours.

2- Garantir une bonne localisation.

3- Assurer que pour un contour il n'y aura qu'une détection.

La figure 3.7 présente un exemple d'application de la détection de contour par Canny.

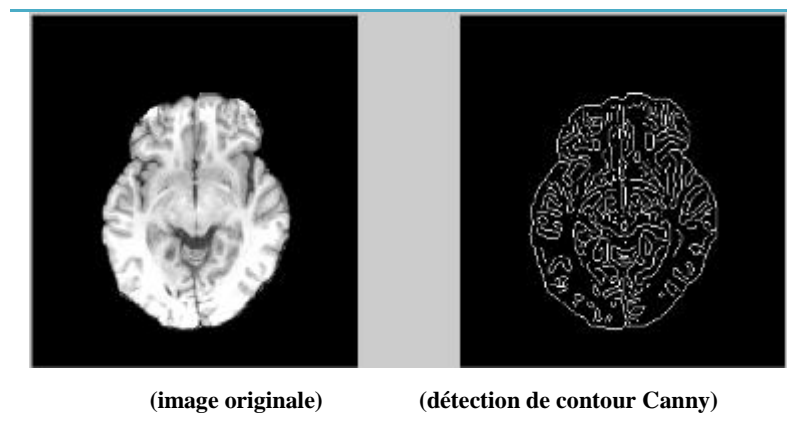


Figure III.10 : Détection de contour par Canny.

III.6.8 Les détecteurs de passage par zéro du Laplacien

Ces méthodes ont été proposées en 1976. Elles ont eu une grande importance historique, étant considérées comme le prototype du détecteur de contour inspiré des systèmes biologiques. Elles utilisent le fait que le passage par zéro du Laplacien permet de bien mettre en évidence les extremums de la dérivée. Le passage par zéro du Laplacien correspond en effet bien au maximum du gradient dans la direction du gradient. La figure 3.8 présente un exemple d'application de la détection de contour par le Laplacien.

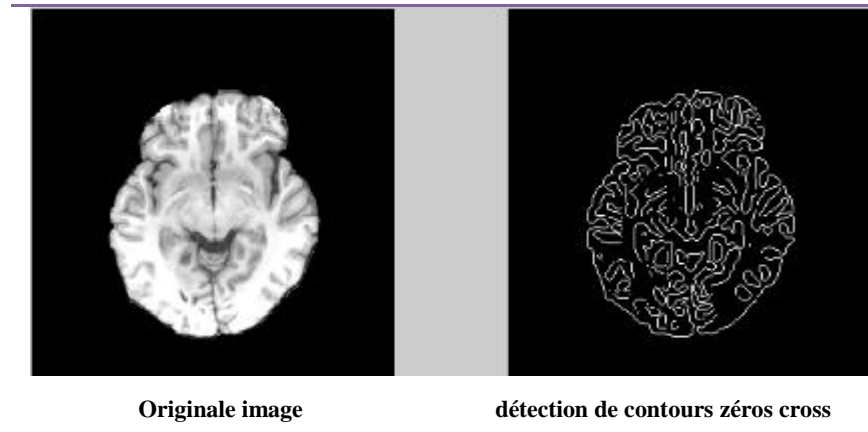


Figure III.11 : Détection de contour passage par zéro.

III.7 Conclusion

Dans ce papier, l'opérateur proposé est à l'œuvre de cette méthodologie qui utilise Simulink le blockset est plus utile pour découverte du Bord de l'Objet à détecter des contours d'un objet, texte de la scène et limites. Vidéos et images de source ouverte peuvent être accompli et donc il peut être rendu effectif facilement et son usage est plus large. Mise en œuvre de ceci la méthodologie qui utilise le blockset Simulink est plus utile.

Ce système fournit pour identifier d'objets et le texte de la scène de la reconnaissance. La performance pour bord la découverte dans une image de la vidéo est évaluée les deux subjectivement et objectivement. Le subjectif l'évaluation de bord a détecté les images de la vidéo montrent cela opérateur proposé, Sobel et Prewitt et Prudent l'objet exposé de l'opérateur meilleures performances respectivement.

Chapitre IV

























Expérimentation & validation

IV.1 Introduction

Nous avons présentés les résultats des différents operateurs suivant tel que : Sobel, Prewitt, Robert, et Canny,...etc.

En présente les valeurs des critères d'évaluations, que nous avons fixés dans chaque cas un operateur parmi les operateurs utilises dans notre travail. Les résultats obtenus pour les différents opérateurs sont présents dans les figures suivantes.

Tableau IV.1 : représenté les résultats des opérateurs utilisés.

L'image	1600	1601	1602	1605
L'operateur				
Original color image				
Gray-level image				
Robert Détection				
Sobel Détection				
Prewitt Détection				
Canny Détection				

Que nous avons fixés dans chaque cas un opérateur parmi les opérateurs utilisés dans notre travail. Les résultats obtenus pour les différents opérateurs sont présents dans les figures suivantes.

Nous avons utilisé l'outil MATLAB, pour traiter l'image et appliquer les différents opérateurs de détection de texte dans l'image, les résultats de traitement présenté dans la figure IV.1 c'est l'opérateur de Sobel avec histogramme de l'image de Sobel présenté dans la figure IV.2.



Figure IV.1 : (a) Image originale, (b) l'image détectée par Sobel.

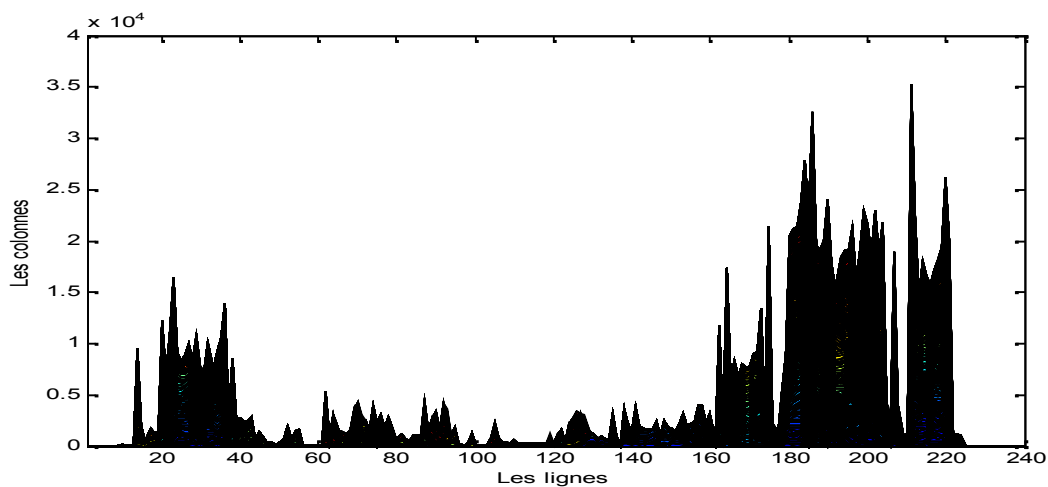


Figure IV.2 : Histogramme de l'image de l'image (b)

Dans la figure IV.3 c'est l'opérateur de Prewitt avec histogramme de l'image de Prewitt présenté dans la figure IV.4.



(c)



(d)

Figure IV.3 : (c) Image original, (d) Image détection de Prewitt.

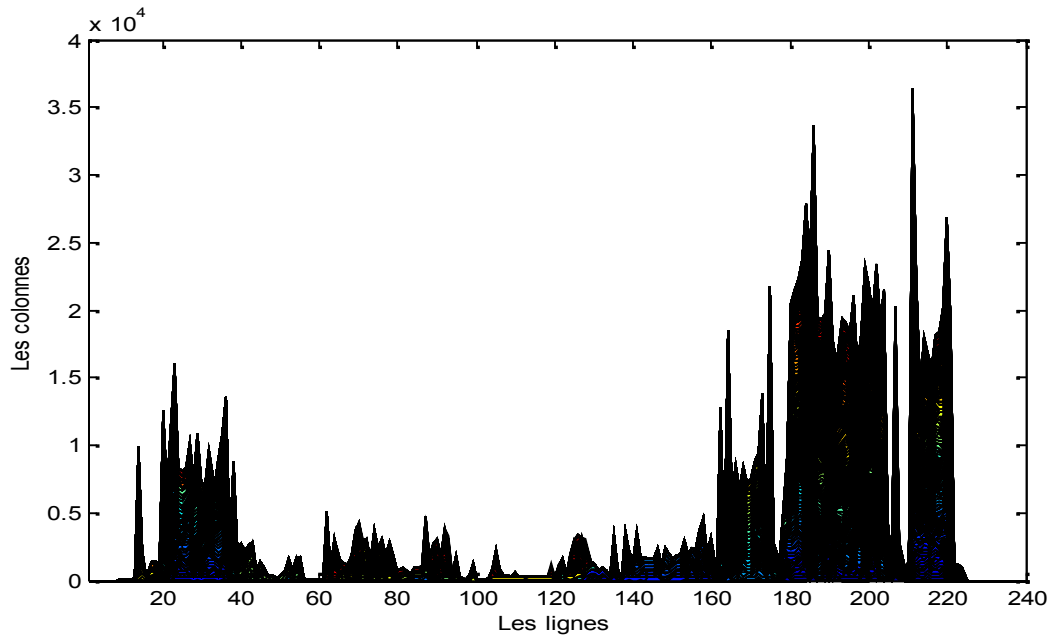


Figure IV.4 : Histogramme de l'image (d).

Dans la figure IV.5 c'est l'opérateur de Canny avec histogramme de limage de Canny présenté dans la figure IV.6.



(e)



(f)

Figure IV.5 : (e) Image original, (f) Image détection de Canny.

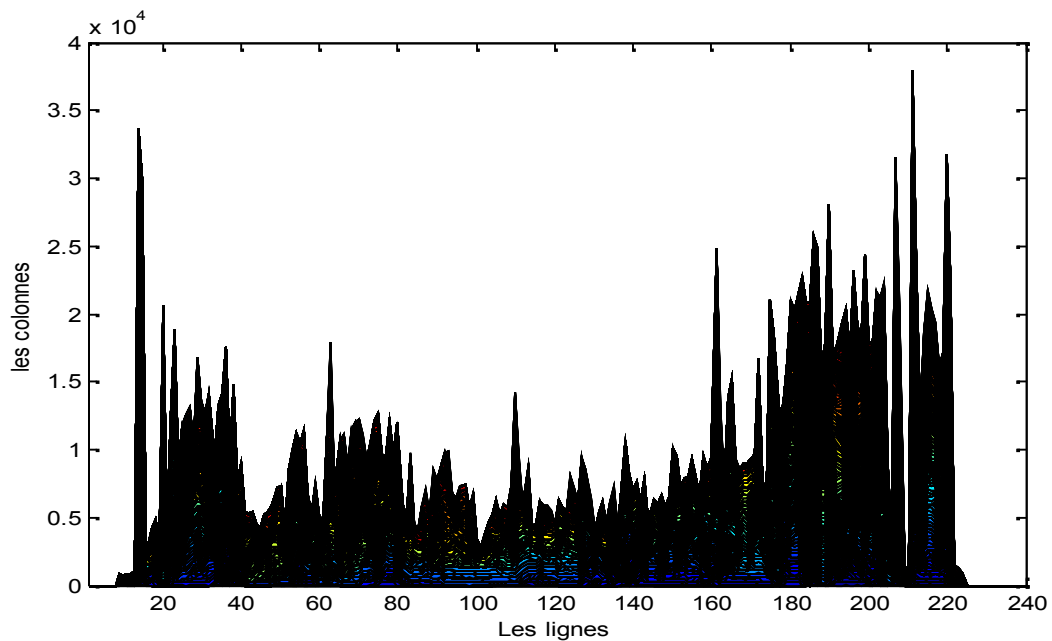


Figure IV.6 : Histogramme de l'image (f).

Dans la figure IV.7 c'est l'opérateur de Robert avec histogramme de l'image de Robert présenté dans la figure IV.8.



(g)



(h)

Figure IV.7 : (g) Image original, (h) Image détection de Robert.

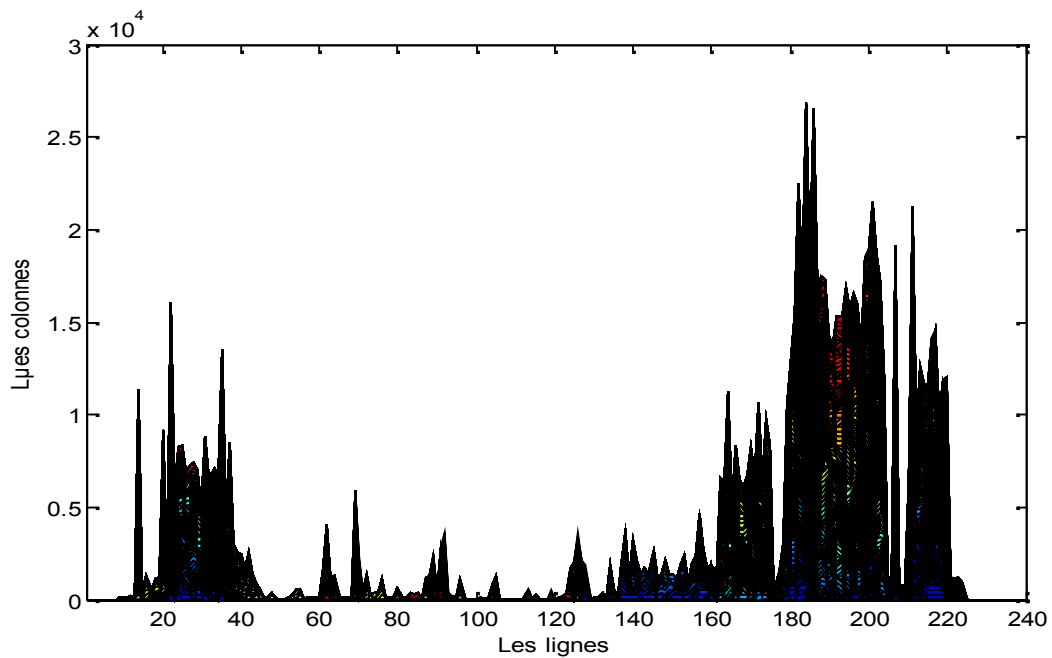


Figure IV.8 : Histogramme de l'image (h).

Dans la figure IV.9 c'est l'image en Niveau de gray avec histogramme présenté dans la figure IV.10.



Figure IV.9 : (i) Image original, (j) Image en Niveau de gray.

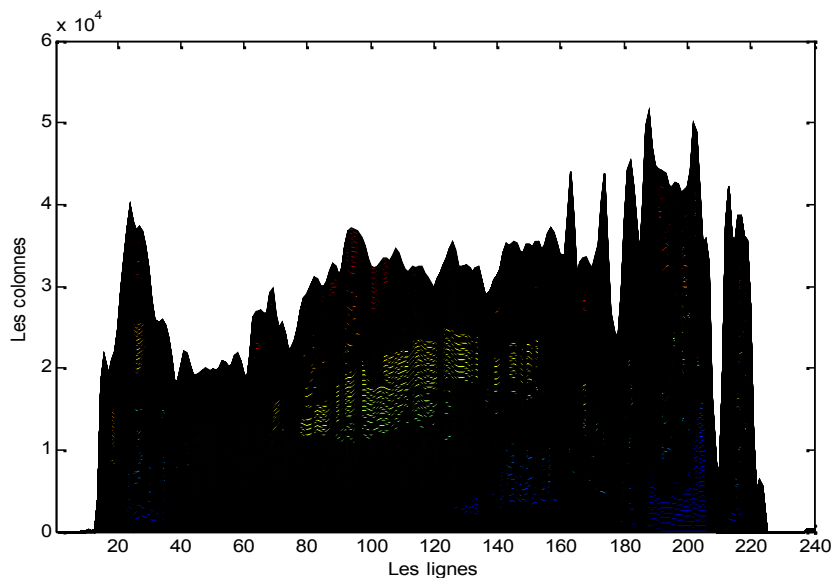


Figure IV.10 : Histogramme de l'image de l'image (j).

IV.2 Résultat des critères

Maintenant nous allons affichés ici les valeurs des critères d'évaluations, que nous avons fixés dans chaque cas un operateur parmi les operateurs utilises dans

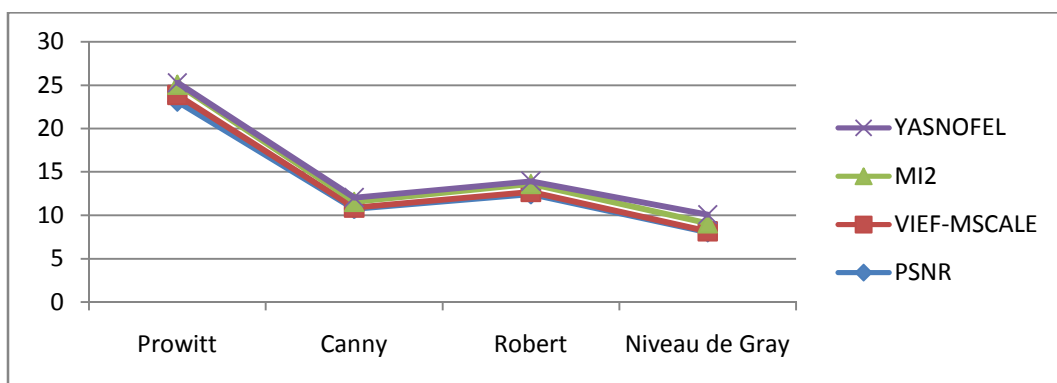
notre travail. Les résultats obtenus pour les différents opérateurs sont présents dans les figures suivantes.

IV.2.1 Nous avons fixe Sobel

Dans ce premier cas nous allons considérer que l'opérateur de Sobel comme une image de référence et nous calculons les différents critères comme démontre le tableau de la figure IV.11.

Critères →	PSNR	VIFP_MSCALE	MI2	YASNOFFL
Opérateur ↓				
PREWITT	23.0710	0.7648	1.2166	0.1924
CANNY	10.7251	0.1791	0.9364	0.4890
ROBERT	12.4206	0.2625	0.9214	0.3040
NIVEAU DE GRAY	8.0414	0.1191	0.9214	0.9798

(a)



(b)

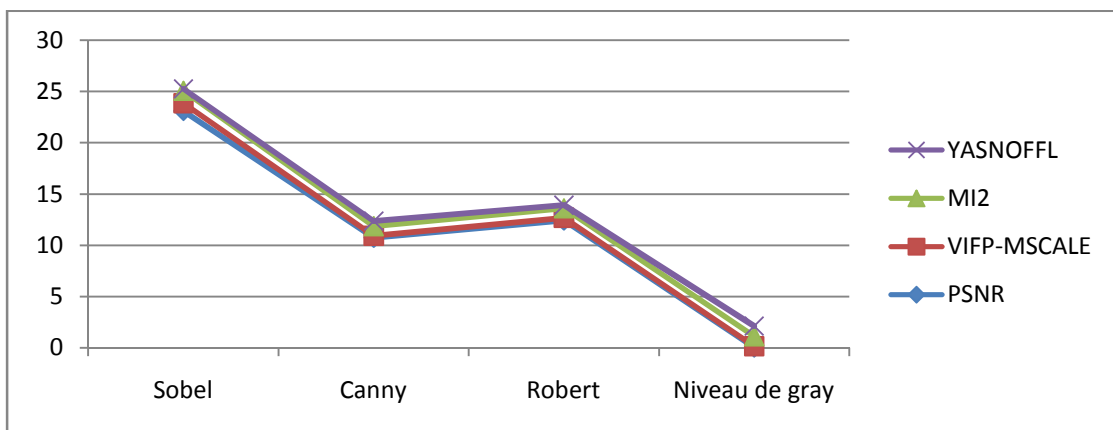
Figure IV.11 : (a) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (b) est l'histogramme de tableau.

IV.2.2 Nous avons fixe l'image Prewitt

Dans ce deuxième cas nous allons considérer que l'opérateur de Prewitt comme une image de référence et nous calculons les différents critères comme démontre le tableau de la figure IV.12.

Critères →	PSNR	VIFP_MSCALE	MI2	YASNOFF
Opérateur ↓				
SOBEL	23.0710	0.7648	1.2166	0.1924
CANNY	10.7473	0.1784	0.9365	0.4889
ROBERT	12.4172	0.2622	0.9211	0.3040
NIVEAU DE GRAY	8.0385	0.1169	0.9793	0.9817

(c)



(d)

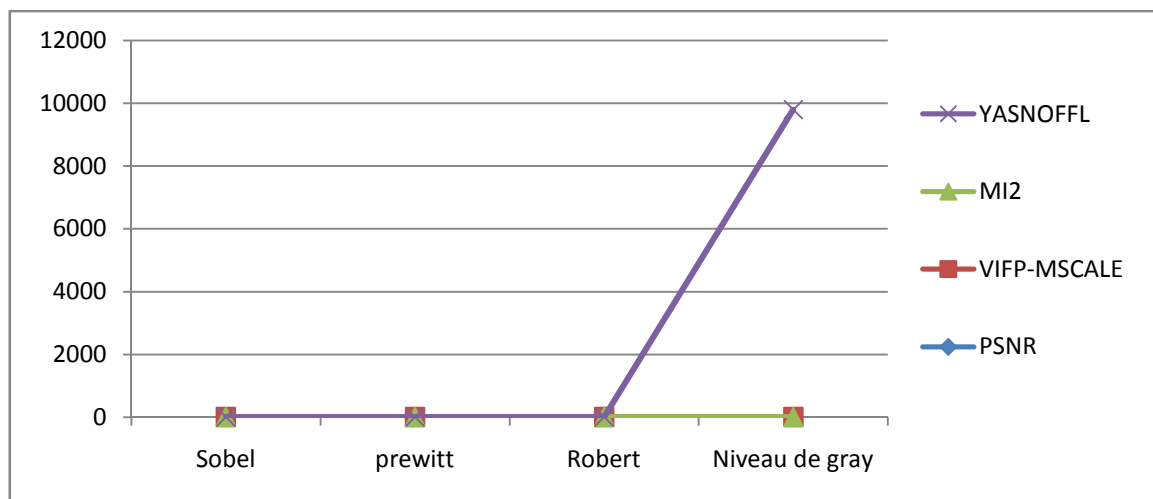
Figure IV.12 : (c) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (d) est l'histogramme de tableau.

IV.2.3 Nous avons fixe l'image Canny

Dans ce premier cas nous allons considérer que l'opérateur de Canny comme une image de référence et nous calculons les différents critères comme démontre le tableau de la figure IV.13.

Critères →	PSNR	VIFP_MSCALE	MI2	YASNOFF
Opérateur ↓				
SOBEL	10.7251	0.1791	0.9364	0.4890
Prewitt	10.7473	0.1784	0.9365	0.4889
ROBERT	9.8466	0.0747	0.9190	0.4889
NIVEAU DE GRAY	7.7612	0.0943	0.9838	0.9781

(e)



(f)

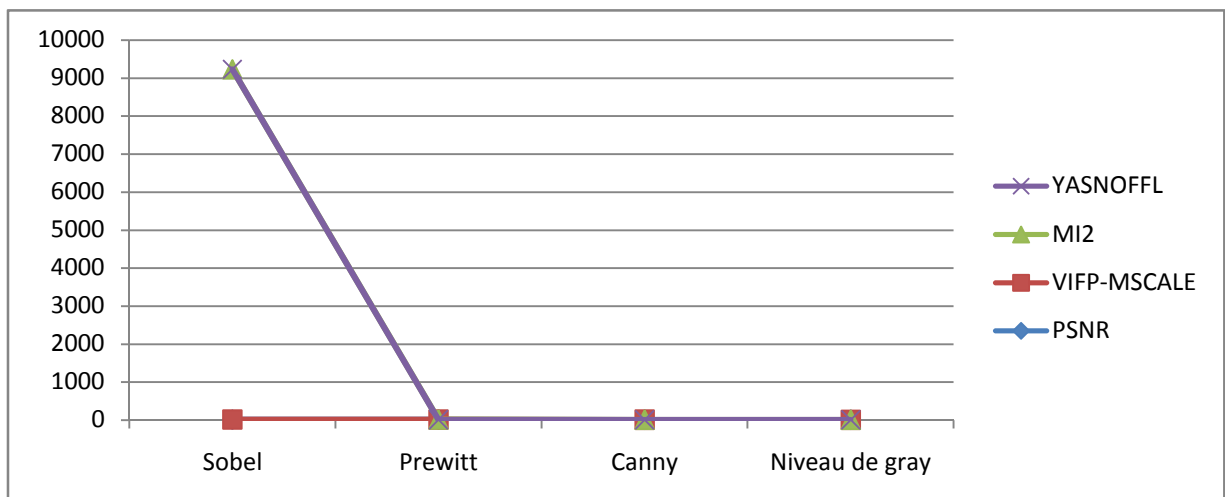
Figure IV.13 : (e) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (f) est l'histogramme de tableau.

IV.2.4 Nous avons fixes l'image Robert

Dans ce premier cas nous allons considérer que l'opérateur de Robert comme une image de référence et nous calculons les différents critères comme démontre le tableau de la figure IV.14.

Critères →	PSNR	VIFP_MSCALE	MI2	YASNOFF
Opérateur ↓				
SOBEL	12.4206	0.2625	0.9214	0.3040
PREWITT	12.4172	0.2622	0.9211	0.3040
CANNY	9.8466	0.0747	0.9190	0.4889
NIVEAU DE GRAY	8.1227	0.11 41	0.9743	0.9814

(g)



(h)

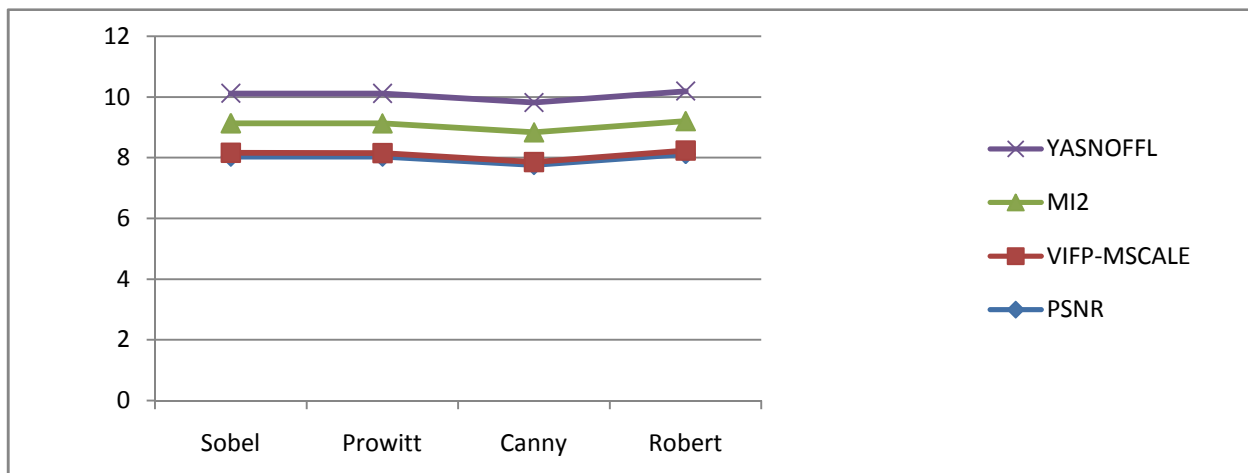
Figure IV.14 : (g) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (h) est l'histogramme de tableau.

IV.2.4 Nous avons fixe l'image de Niveau de gray

Dans ce premier cas nous allons considérer que l'image de niveau de gray comme une image de référence et nous calculons les différents critères comme démontre le tableau de la figure IV.15.

Critères →	PSNR	VIF P_MSCALE	MI2	YASNOFF
Opérateur ↓				
SOBEL	8.0414	0.1191	0.9797	0.9798
PREWITT	8.0385	0.1169	0.9793	0.9817
CANNY	7.7612	0.0943	0.9838	0.9781
ROBERT	8.1227	0.11 41	0.9743	0.9814

(i)



(j)

Figure IV.15 : (i) est les Valeurs des critères pour chacun des 4 résultats de segmentation de l'image, (j) est l'histogramme de tableau.

IV.3 Conclusion

L'expérience réalisée sur plusieurs images et sur vidéos et les conclusions positives issues de la comparaison avec les résultats d'autres méthodes ont permis une première validation de notre approche : Nous avons présenté dans ce chapitre les algorithmes utilisés pour segmenter une image vidéos ainsi les critères de validation.

Les résultats obtenus démontrent que le critère de Yasnoff est meilleur dans tous les cas que nous avons appliqué ces critères.

Conclusion générale

Conclusion générale

Ce mémoire a pour but est de localiser toutes les zones de texte artificiel dans des images extraites de vidéos. Le texte artificiel est destiné à être lu par un observateur contrairement aux autres types de texte que l'on peut rencontrer dans une scène qui ne sont pas spécialement destinés à être lus. Ainsi, les textes artificiels sont des zones d'image horizontales, non occultées, fortement contrastées et assez grandes. Nous avons présenté une méthode d'extraction de texte qui combine l'utilisation des opérateurs morphologiques et des critères de cohérence spatiale adaptables à toutes les régions dans diverses orientations.

L'approche présentée dans ce mémoire est une contribution au développement d'outils pour le contenu des images animées.

En effet, la détection, la segmentation et la reconnaissance du texte sont des tâches essentielles dans plusieurs applications telles que la segmentation des journaux télévisés.

Bibliographie

- [1] **F. M. Blondel.** "Acquisition, traitement et restitution des images numériques", à l'occasion des Cinquièmes Journées des Sciences Physiques en 1992.
- [2] **Hadjila F. & Bouabdellah R.**, « Reconnaissance des visages par les réseaux de neurones », *Mémoire de PFE, Université de Tlemcen, 2003.*
- [3] **Naciri H., Chaoui N.** « Conception et Réalisation d'un système automatique d'identification des empreintes digitales », *Mémoire de PFE, Université de Tlemcen, 2003.*
- [4] **Hadjila F. & Bouabdellah R.**, « Reconnaissance des visages par les réseaux de neurones », *Mémoire de PFE, Université de Tlemcen, 2003.*
- [5] **Ariane Herbulot.** *Automatique traitement du signal et des images "Mesures statistiques non-paramétriques pour la segmentation d'images et de vidéos".figure [I.11].*
- [6] *Traitement d'image de l'adresse du site suivant : ' <http://ow.ly/URBHf> '. Figure [I.13]*
- [7] *Traitement d'image de l'adresse du site suivant : ' <http://ow.ly/URBHf> '. Figure [I.14]*
- [8] **A. Gagalowic**, *de cours dispensés à l'ESIEA, l'année 2002.*
- [9] **E. TISSERAND, J. PAUTTEX, P. SCHWEITZER**, *Analyse et traitement des signaux, Méthodes et applications au son et à l'image, Dunod, Paris, 2008.*
- [10] **MURIEL GASTAUD.** *Segmentation D'image Par Une approche Contour Actifs*

-
- [11] **C. VACHIER et F. MEYER.** *THE VISCOUS WATERSHED transforms.* *Journal of Mathematical Image and Vision*, 22(2-3): 251-267, 1005.
- [12] **S. BEUCHER et F. MEYER.** *THE MORFOLOGICAL APROGH TO SEGMENTAION: The watershed transformation.* *Mathematical Morphology in Image Processing.*
- [13] **Christopher Zach, David Gallup and JM Frahm.** *Fast gain-adaptive KLT tracking on the GPU.* In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2008.
- [14] **Brox and Jitendra Malik.** *Object segmentation by long term analysis of point trajectories.* In *European Conference on Computer Vision.* Springer, 2010.
- [15] **R. Deriche.** *Using Canny's criteria to derive a recursively implemented optimal edge detector.* *International Journal of Computer Vision*, vol. 1, no. 2, pages 167–187, 1987.
- [16] **Jianming Zhang, LL Presti and Stan Sclaroff.** *Online multi-person tracking by tracker hierarchy.* In *Conference on Advanced Video and Signal-Based Surveillance*, 2012. LL25 Jun 2008
- [17] **S. Sastry O. Shakernia, R. Vidal.** *Omnidirectional egomotion estimation from back-projection flow.* *Madison Wisconsin, USA, June 2003.*
- [18] **Chia-Hung Wei,** *Computer Methods and Programs in Biomedicine Volume 106, Issue 3, Pages 234-248, June 2012*
- [19] **D Mbark Iggane, Driss Mammass, Ali Khenchaf and Jean-Christophe**

Segmentation d'images sonar par matrice de cooccurrence. International Conference: Sciences of Electronic, Technologies of Information and Telecommunications 2009 –TUNISIA.

[20] *An Adaptive **K-means** Clustering Algorithm for Breast Image Segmentation, Bhagwati Charan, Patel, November 2010*

[21] *Thomas Brox^{1,2} and Jitendra Malik. "Object Segmentation by Long Term Analysis of point Trajectories". University of California at Bermany.*

[22] *AR Zamir, Afshin Dehghan and Mubarak Shah. GMCP-Tracker: Global*

[23] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.

[24] D. Coquin, P. Bolon, and B. Ionescu. Dissimilarity measures in color spaces. In *16th ICPR*, volume 1, 2002.

[25] P. Felzenszwalb and D. Huttenlocher. Efficiently computing a good segmentation. In *Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, 1998.

[26] L. Guigues. Image segmentation comparison using hierarchical model for nm region matching. In *Proc. of 2nd IAPR TC-15 Workshop on Graph-based Representations in Pattern Recognition*, Haindorf, Austria, 1999.

[27] S. Chabrier, C. Rosenberger, H. Laurent, B. Emile, and P. March. Evaluating the segmentation result of a gray-level image. In *12th EUSIPCO*, Vienne, Austria, september 2004.

[28] V. Chalana and Y. Kim. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans. on Medical Imaging*, 16(5):642–652, 1997.

[29] J.P. Cocquerez and S. Philipp editors. *Analyse d'images: filtrage et segmentation*. Masson, Paris, 1995.

-
- [30] L. Guigues. Modèles multi-échelles pour la segmentation d'images. PhD thesis, Université de Cergy-Pontoise, 2003.
- [31] J. S. Cardoso and L. Corto-Real. Toward a generic evaluation of image segmentation. *IEEE trans. on Image Processing*, 14(11), 2005.
- [32] S. Yu and J. Shi. Segmentation given partial grouping constraints. *IEEE trans. on Image Processing*, pages 173–183, 2004.
- [33] D. R. Martin. An empirical approach to grouping and segmentation. PhD thesis, University of California, Berkeley, USA, 2002.
- [34] W. A. Yasnoff, W. Galbraith, and J.W. Bacus. Errormeasures for objective assessment of scene segmentation algorithms. *AQC*, 1:107–121, 1979.
- [35] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.
- [36] **Christian Wolf**. Thèse « Détection de textes dans des images issues d'un flux Vidéo pour l'indexation sémantique », Spécialité Informatique, Année 2013
- [37] Nick Kanopoulos, et.al. ; “Design of an Image Edge Detection Filter using the Sobel Operator”, *Journal of Solid State Circuits, IEEE*, vol. 23, Issue: 2, pp. 358-367, April 1988.
- [38] Seif, A.,et.al. ; “A hardware architecture of Prewitt edge detection”, *Sustainable Utilization and Development in Engineering and Technology (STUDENT)*, 2010 *IEEE Conference, Malaysia*, pp. 99 – 101, 20-21 Nov. 2010.
- [39] J.-P. COCQUEREZ et S. PHILIP (1995) « Analyse d'images : filtrage et segmentation »

Résumé

Les systèmes d'indexation ou de recherche par le contenu disponibles actuellement travaillent sans connaissance (systèmes pré attentifs). Malheureusement les requêtes construites ne correspondent pas toujours aux résultats obtenus par un humain qui

Interprète le contenu du document. Le texte présent dans les vidéos représente une caractéristique à la fois riche en information et cependant simple, cela permet de compléter les requêtes classiques par des mots clefs.

Nous présentons dans cet article un projet visant à la détection et la reconnaissance du texte présent dans des images ou des séquences vidéo. Nous proposons un schéma de détection s'appuyant sur la mesure du gradient directionnel cumulé. Dans le cas des séquences vidéo, nous introduisons un processus de fiabilisation des détections et l'amélioration des textes détectés par un suivi et une intégration temporelle.

Abstract

Systems of indexing or research by the available content work currently without knowledge (systems attentive meadow). Unfortunately the constructed requests correspond not always to results gotten by a human who Interpret the content of the document. The present text in videos represents a characteristic at a time rich in information and however simple, it permits to complete the classic requests by words keys.

We present in this article a project aiming to the detection and the recognition of the present text in pictures or the sequences video. We propose a diagram of detection leaning on the measure of the directional gradient accumulated. In the case of the sequences video we introduce a process of detection fiabilisation and the improvement of texts detected by a follow-up and a temporal integration.