RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ SAAD DAHLEB DE BLIDA 1
FACULTÉ DES SCIENCES
DÉPARTEMENT D'INFORMATIQUE

MÉMOIRE DE MASTER

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Génie des Systèmes Informatiques

Intitulé

Segmentation d'une vidéo en plans

Présenté par : **BENAROUS Amina**
DJOUAH Hamida

Soutenu le : 20/09/2015

Promoteur : Mr **CHERIF-ZAHAR**

Devant le jury composé de :

Président : GUESSOUM D.	MAB	USDB
Examinateur1 : KEMACHE A.	MAB	USDB
Examinateur2 : NEHAL D.	MAB	USDB

Promotion
2014-2015

Remerciement

Nous glorifions et remercions notre Dieu, Allah le Tout Puissant, Créateur des cieux et de la terre, qui nous a accordé le souffle de vie pour accomplir et pouvoir présenter ce modeste travail, que l'honneur et la gloire Lui soient rendus.

En ce moment qui marque la fin de nos études un mot de remerciement et de reconnaissance mérite d'être prononcé.

Nos plus profonds remerciements s'adressent à Mr ***CHERIF-ZAHAR*** *pour son encadrement, son soutien, ses conseils et ses orientations tout au long de ce travail. Nous la remercions vivement.*

Nous tenons également à remercier tous les membres du jury pour avoir accepté de consacrer une partie de leur temps à la lecture de ce modeste travail, de l'évaluer et le discuter.

Nous remercions très sincèrement tous ceux qui de loin ou de près d'une manière ou d'une autre ont contribué à la réalisation de ce mémoire.

Nous remercions tous les membres du département Informatique dirigé par le Chef du Département Mme ***BENSTTITI.***

Résumé

Lorsqu'un réalisateur d'un montage vidéo fait son travail, il colle des plans différents qu'il serait très fastidieux à un humain de retrouver car nécessitant de parcourir toute vidéo et repérer manuellement tous les changements de plans.

Hors, pour diverses raisons, il serait souhaitable de retrouver les différents plans d'une vidéo de façon automatique et de la façon la plus sûre possible ouvrant ainsi le champ à bien des applications telles indexation de vidéos, élimination de scènes indésirables, repérages de scènes particulières etc.

Le présent travail se propose de faire un défrichage de ce domaine de l'informatique appliquée et d'explorer autant que possible les différentes méthodes qui ont vu le jour et tentant d'en expérimenter quelques-unes.

***Mots clés** : Vidéo, Segmentation temporelle, Changement de Plan, Descripteurs Visuels, Histogramme de couleur, Détection du mouvement*

Abstract

When a director of video editing done his job correctely, it sticks different plans. it would be very tedious to find them by a human because that are nessacery to requiring to browse all the video and identify any changes of cut manually.

for various reasons , it would be desirable to find the different cuts of video automatically and the safest possible way and opening the field in many applications such as indexing videos, removing unwanted scenes , scenes of special trails etc..

This work intends to make a clearing this area of applied computing, and explore as much as possible the different methods that have emerged and tempting to try a few one.

***Keywords**: Video, temporal segmentation, Moving Up, Visual Descriptors, color histogram, Motion Detection*

ملخص

عندما يكون محرر الفيديو يقوم بعمله, يلصق مقاطع مختلفة التي تصعب علينا العثور عليها وذلك لتطلبها تغطية الفيديو كاملا و العثور عليها يدويا

قد يكون من المرغوب فيه العثور على هذه المقاطع المختلفة من الفيديو تلقائيا وبأسلم طريقة ممكنة هذا يفتح لنا مجال واسع في العديد من التطبيقات مثل فهرسة الفيديوهات, و إزالة المشاهد الغير المرغوب فيها أو تعديلها...إلخ

يهدف هذا العمل الى تعزيز مجال الحاسوب التطبيقي و استكشاف القدر الممكن من مختلف الطرق الموجودة الى يومنا هذا و تطبيق البعض منها.

كلمات البحث: فيديو، تجزئة الزمنية، تغيير المقطع، واصفات بصرية، الرسم البياني للون، وكشف الحركة

Glossaire

3D : Trois dimensions ou tridimensionnel ou 3D sont des expressions qui caractérisent l'espace qui nous entoure, tel que perçu par notre vision, en termes de largeur, hauteur et profondeur.

2D : Deux dimensions, largeur et hauteur,

Annotation : faire des remarques (des notes) sur un média pour l'expliquer ou le commenter.

AVI (*Audio Video Interleave : Video For Windows*) L'Audio Video Interleave (audio vidéo entrelacée) est un format de fichier conçu pour stocker des données audio et vidéo.

Bijective : à la fois injective et surjective qui établit entre les éléments de deux ensembles une correspondance telle que tout élément de l'un a un correspondant et un seul dans l'autre.

Bruit momentanée : est la présence d'informations parasites qui s'ajoutent de façon aléatoire aux détails de la scène.

CIELab : L'espace Lab, appelé également CIELab, a été introduit par la Commission Internationale d'Éclairage (CIE) en 1976. Sa propriété principale est son uniformité comparée à l'espace RGB.

CMY : est un espace de couleur directement déduit de l'espace RGB. Il est souvent utilisé par les imprimantes couleur. Les composantes de cet espace sont C pour Cyan, M pour Magenta et Y pour Jaune *(Yellow)*.

Corrélogramme : est une représentation graphique mettant en évidence une ou plusieurs corrélations entre des séries de données. Et permet de visualiser des données sous différentes formes.

Débit binaire est une mesure de la quantité de données numériques transmises par unité de temps. Il est le plus souvent exprimé en bits par seconde.

Ensemble flou : est une théorie mathématique du domaine de l'**algèbre abstraite**. Elle a été développée par Lotfi Zadeh en 1965 afin de représenter mathématiquement l'imprécision relative à certaines classes d'objets et sert de fondement à la logique floue.

Entrelacement : (en anglais *interlace*) : ou **balayage entrelacé**, est une technique destinée à doubler le nombre d'images par seconde perçu sans augmenter le débit.

Entropie : L'entropie est une fonction mathematique permettant de mesurer la quantite d'information contenue dans une source d'information (i.e. dans une VA). Elle peut se definir comme une mesure de desordre et d'incertitude de l'information.

Espace métrique : est un ensemble x muni d'une distance d, ses éléments sont habituellement appelés des points.

Frame : Les frames sont les images qui constituent la vidéo telle que chaque frame présente un état instantané. Elle est perceptible par l'œil humain.

Frame clé : Un ***frame clé*** est une image qui exprime bien le contenu d'un plan, donc un plan peut comporter une ou plusieurs frames selon la taille ou la complexité du plan.

Histogramme : Un *histogramme* représente le mode de répartition des pixels dans une image en traçant le nombre de pixels (graphe) correspondant à chaque niveau d'intensité de la couleur.

HSV : C'est un espace dérivé de l'espace RGB, le plus souvent, utilisé dans des applications informatiques de graphisme. Les couleurs dans cet espace sont représentées selon des notions de teinte *(Hue)*, de pureté *(Saturation)* et de luminosité *(Value)*.

Inégalité triangulaire : exprime l'idée que la distance est une mesure minimale. Cela se traduit par le fait que la ligne droite est le chemin le plus court.

Granularité : définit la taille du plus petit élément, de la plus grande finesse d'un système. Quand on arrive au niveau de granularité d'un système, on ne peut plus découper l'information.

Loi de X^2 (Chi deux) : est une loi à densité de probabilité

Longueur d'onde : une perturbation qui se propage dans un milieu sans modifier de façon permanente ses propriétés.

Matrice : Une matrice à m lignes et n colonnes est un tableau rectangulaire de mn nombres, rangés ligne par ligne. Il y a m lignes, et dans chaque ligne n nombres.

Multidimentionnel : Qui a des dimensions multiples, qui concerne des niveaux variés.

Méthode OTSU : Le nom de cette méthode provient du nom de son initiateur, *Nobuyuki Otsu.*

Pixel : (souvent abrégé **px**) est l'unité de base permettant de mesurer la définition d'une image numérique matricielle.

Précision mesure la proportion de documents pertinents retrouvés parmi l'ensemble des documents retrouvés.

Rappel mesure la proportion de documents pertinents retrouvés par rapport au nombre total de documents pertinents dans la collection.

Résumé vidéo : Le résumé est une version courte de la vidéo qui doit contenir l'essentiel de l'information, tout en étant le plus concis possible.

Sommaire

Table des figures

Table des tableaux

Introduction Générale

Aujourd'hui, la vidéo est le document le plus riche, en matière des sens. Elle est composée des seuls deux sens transportables à distance, parmi les cinq avec lesquels l'homme peut communiquer. Elle est utilisée dans des secteurs de plus en plus nombreux de la vie courante et joue un rôle très important dans notre société moderne. Ceci explique le besoin de structurer cette information pour faciliter sa recherche et l'accès à son contenu. Le problème de la consultation des vidéos réside dans la visualisation séquentielle de leur contenu pour pouvoir localiser un passage particulier auquel on s'intéresse.

L'amélioration des services offerts aux utilisateurs des bases de données vidéo doit passer nécessairement par l'introduction de l'accès direct au contenu de ces documents. Il faut alors découper la vidéo en parties (*plans*) significatives auxquelles l'utilisateur aura accès. Les plans représentent l'équivalent d'une table des matières permettant un accès rapide et guidé au contenu du document. Elle constitue une sorte de résumé visuel du document vidéo.

Le découpage de la vidéo en plans (*ou segmentation temporelle*) est une étape cruciale dans des applications telles que la gestion de bases de données multimédia ou la création automatique de résumés de séquences télévisées ou de films. Elle peut être aussi utilisée comme prétraitement pour le suivi d'objet en temps réel dans des scènes dynamiques (acquises avec une caméra en mouvement). Il peut être mené par un opérateur humain, mais seul son automatisation peut constituer une alternative sûre et rapide et ce en palliant aux problèmes inhérents à l'approche manuelle (lenteur, voire impossibilité de traitement lorsqu'il s'agit de base de données volumineuses).

Nous définissons le problème de segmentation vidéo en plan comme la détection automatique des frontières qui séparent les plans.

Notre travail passe en revue les principales méthodes connues à ce jour en matière de segmentation vidéo pour enfin retenir deux méthodes qui semblent intéressantes pour expérimentation. La détection des transitions par la méthode d'histogramme de couleur et par la méthode de mouvement.
L'objectif de ces techniques proposées est de pouvoir détecter avec efficacité les différents types de transitions (*cuts*, *fondu* et *volet*).

Le manuscrit de notre travail est divisé en trois chapitres ; nous commençons par un état de l'art où nous aborderons en premier lieu, les définitions, les notions générales et les descripteurs visuels, ainsi que les distances de similarités entre les descripteurs visuels. En second lieu, nous présenterons les principales de la segmentation vidéo. En se basant sur l'objectif de notre travail, nous allons décrire un panorama des travaux relatifs aux méthodes de la segmentation en plans de la vidéo existante dans la littérature.

Le deuxième chapitre expose les deux méthodes proposées, la méthode d'histogramme de couleur et la méthode de mouvement en détaille avec les algorithmes qui nous avons utilisé dans notre application.

Le troisième chapitre dans le but est de présenter la partie expérimentale et de discuter les différents résultats obtenus par les techniques proposées.

Enfin, nous terminerons par une conclusion générale finalise notre travail qui discute les principaux résultats obtenus et fixe les perspectives envisageables.

Chapitre 1 : Etat de l'art

1.1. Introduction

La segmentation des documents vidéo est une tache incontournable dans le cadre d'un procédé d'indexation audiovisuelle. Différents niveaux de granularité de découpage existent comme par exemple la micro-segmentation et la macro-segmentation [1].

Dans ce chapitre, nous développons l'état de l'art de la segmentation, dans la première section nous définissons des termes qui servent comme notions clés pour notre travail. Dans la deuxième section, nous donnons une vue générale sur la segmentation temporelle qui est obtenue par détection des changements de plans. Après nous présentons un panorama des méthodes de détection des changements de plans ainsi qu'une classification des différentes méthodes avec les grands principes de chaque type d'approche.

1.2. Définitions et Notions Générales

1.2.1. Qu'est ce qu'un document vidéo ?

Formellement, un document vidéo est défini comme une combinaison de flux d'information.

Principalement, deux sources d'informations composent ce document à savoir l'image et le son, qui sont synchronisés pour former une histoire.

Le flux visuel comporte une séquence d'images fixes qui selon l'axe temporel apparaissent animées à une fréquence de 24 à 30 images par seconde.

Le flux sonore est composé d'un ou plusieurs canaux (mono, stéréo). Le signal sonore est typiquement échantillonné entre 16000 et 48000 Hertz.

Un troisième flux d'information généralement associée aux documents vidéo est le texte. Il provient soit d'un flux séparé, soit il est dérivé des sources audio et visuelle [2].

Dans le cadre de ce mémoire, le son et le texte n'ont pas été pris en compte et seule l'information apportée par les images a été étudiée.

Pour le stockage, la manipulation et la recherche de documents vidéo, il est nécessaire de se doter d'un moyen d'organiser l'information. On peut considérer qu'il existe plusieurs niveaux de structure liés à la donnée vidéo, il s'agit d'une organisation hiérarchique issue du monde de la production vidéo. Cette hiérarchie met en évidence des séquences de granularités différentes : le document complet, la scène, le plan, puis l'image [2].

Une séquence vidéo brute est une suite d'images fixes, qui peut être caractérisée par trois principaux paramètres :

-*La* ***résolution en luminance*** **:** permet de définir les couleurs possibles pour un pixel. La couleur est codée sur 8 bits pour les niveaux de gris (donc il peut prendre 255 niveaux de gris) et de 24 bits pour les séquences en couleurs.

-*La* ***résolution spatiale*** **:** définit le nombre de lignes et de colonnes de la matrice de pixels.

-*La* ***résolution temporelle :*** est le nombre d'images par seconde.

La valeur de ces trois paramètres détermine l'espace mémoire nécessaire pour stocker une séquence vidéo [3].

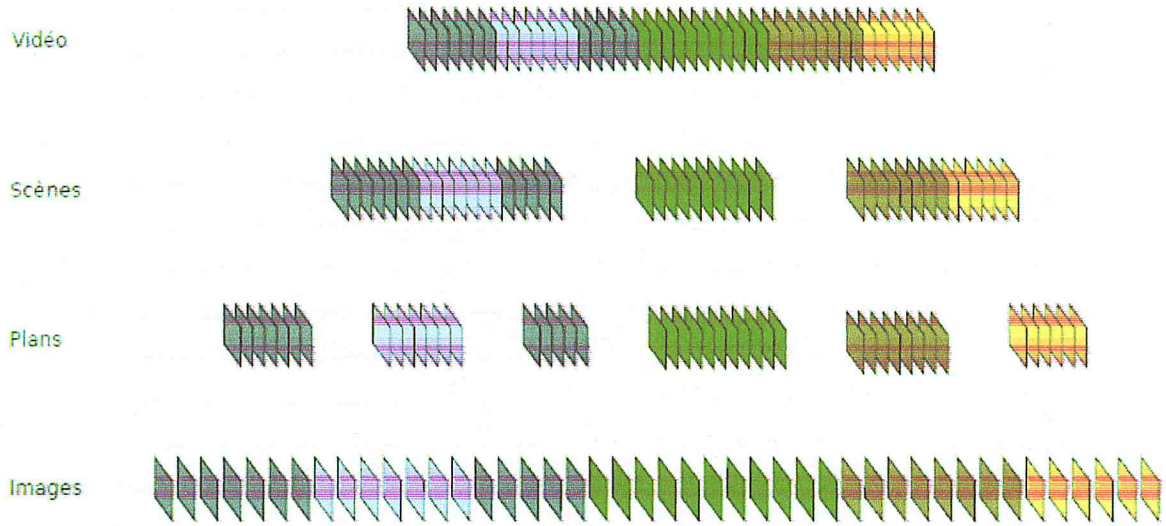


Figure 1 – Différents niveaux de structuration d'une vidéo [2]

1.2.2. L'unité physique : le Plan

Définition d'un plan : *Le plan correspond à l'unité audiovisuelle élémentaire. Il peut être défini physiquement comme la plus petite unité d'une continuité d'un film ou d'un produit vidéo sans coupure de caméra ou de raccord"*[1].

Un **plan** est défini comme une suite d'images dont le contenu est homogène (Figure 2) issues d'une acquisition continue d'une même caméra donnée. Le plan est souvent l'unité temporelle la plus petite pour une séquence vidéo c.-à-d. le plus élémentaire dans l'opération de montage d'une vidéo. Ce montage consiste, pour simplifier, à concaténer des extraits (plans) de bande vidéo ou de film que le réalisateur aura retenu, donc une vidéo représentée par un ensemble de ses plans [4].

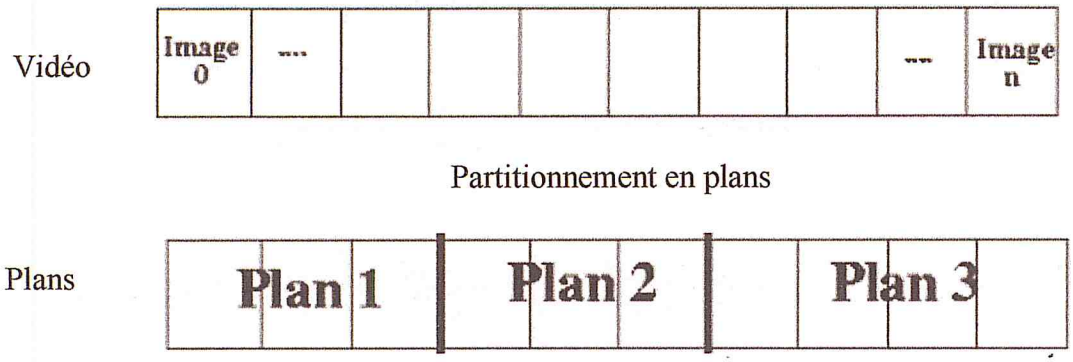


Figure 2 : Structure interne de la vidéo [4]

Montage vidéo et raccord vidéo

Lors du montage d'une vidéo, le raccord entre deux plans peut se faire par un raccord vidéo ou par un collage physique de la bande du film. Ces manipulations entraînent parfois une altération des images proches du raccord, qui peut perturber les étapes d'analyse du contenu de la vidéo [4].

Le passage d'un plan au plan suivant (raccord vidéo ou transition) peut être considéré comme une ponctuation du document visuel. Les réalisateurs ont cherché à exploiter ces transitions en diversifiant leur nature "technique" pour exprimer différents messages de relation entre plans. On peut les classer ces techniques en deux classes [4]:

1.2.2.1. La transition brusque (appelée *Cut*)

C'est le raccord le plus simple et le plus fréquent entre deux plans, la dernière image du premier plan est suivi par la première image du second plan c.-à-d. un changement de situation radicale où il n'y a aucune relation entre deux plans. Aucun effet n'est inséré entre les deux plans, comme le montre la Figure 3.

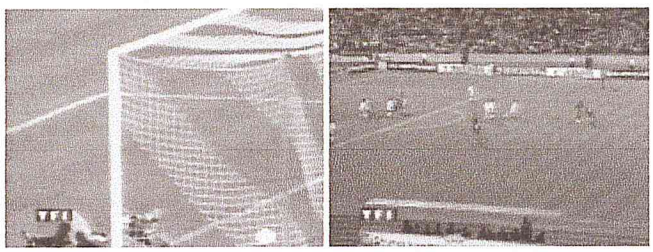


Figure 3 : Exemple d'une transition brusque. (Transition Coupure (*cut*) [5]

1.2.2.2. La transition progressive

Dans le cas où les deux plans sont connectés en utilisant un effet particulier, on parle de transition progressive. Différents types de transitions peuvent être utilisés. Les plus connus sont le *fondu* et le *volet*.

1.2.2.3. Le Fondu

D'après [5], on distingue le fondu du noir vers un plan, d'un plan vers le noir, ou d'un plan vers un autre plan. Au cours d'un fondu, le niveau de chaque pixel des images intermédiaires (appartenant à la transition progressive) est calculé en fonction des niveaux des pixels de la dernière image du premier plan et de la première image du second plan. La proportion varie au cours de la transition de 0 à 1 pour la première image du second plan et de 1 à 0 pour la dernière image du premier plan. (Figure 4)

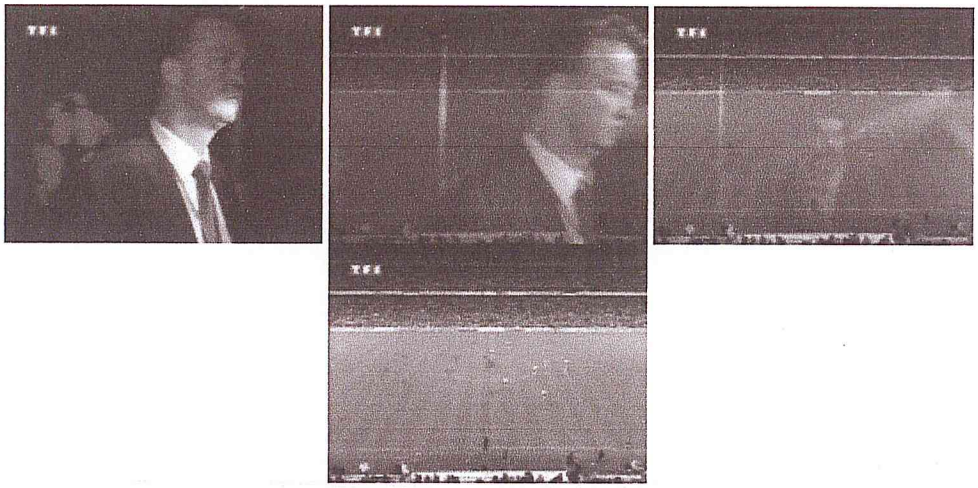


Figure 4 : Exemple d'une transition progressive de type fondu [5].

1.2.2.4. Le volet

Le *volet* consiste a passer progressivement d'un plan I (t) 1 à un autre plan I (t) 2 en variant dans le temps la proportion en terme d'espace dans l'image allouée à chaque plan (voir figure 5). Il peut être effectué selon un déplacement horizontal, vertical ou plus complexe (rotation…). Cette transition peut par exemple produire un effet de page que l'on tourne, ce qui peut expliquer son utilisation abondante dans les journaux télévisés [4].

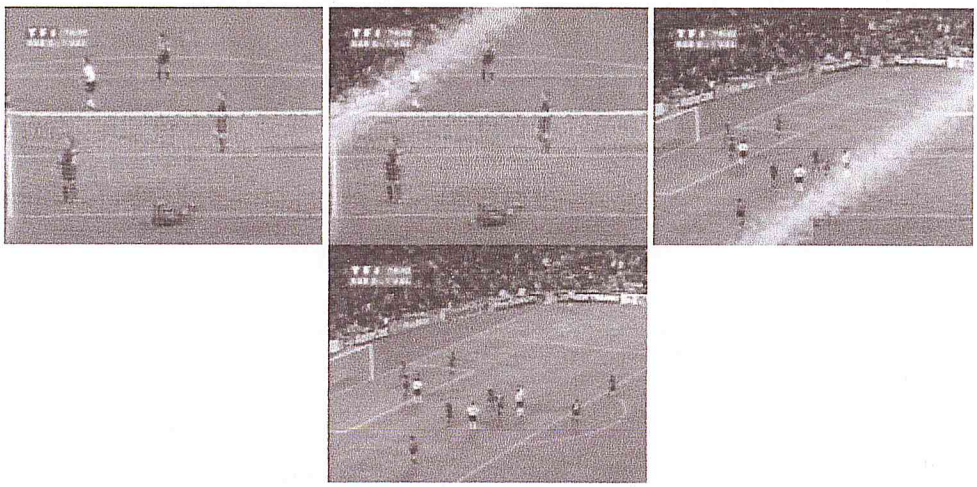


Figure 5 : Exemple d'une transition progressive de type volet [5].

1.2.3. L'unité sémantique : la scène

Une ***scène*** peut être vue comme un ensemble de plans successivement qui partagent le même contenu en termes d'actions (comme le mouvement, la ressemblance visuelle), de lieu et de temps (comme ville/montagne/extérieur/foret,environnement extérieur/intérieur, jour/nuit) [6]

1.2.4. Les images de références

L'image de référence, est une image exploitée pour prédire une image et estimer le mouvement. Cette image peut être simplement une image précédemment codée dans la séquence [6].

1.2.5. Extraction des caractéristiques

Les caractéristiques sont considérées comme un outil pour décrire le contenu d'une vidéo. Elles sont les informations que l'on extrait de la vidéo de telle sorte qu'elles représentent d'une manière appropriée son contenu. Elles sont conservées dans un vecteur index, elles sont classifié en deux groupes : ***caractéristique de bas niveau ou brutes***, et ***caractéristiques de haut niveau ou sémantiques*** selon leur complexité et utilisation des sémantiques [7].

1.2.5.1. Caractéristiques de bas niveau

Généralement les caractéristiques de bas niveau les plus utilisées pour décrire la vidéo sont : la couleur, la texture, la forme et le mouvement.

a) La couleur

La couleur représente la caractéristique la plus couramment utilisée pour la représentation des images. Elle est inchangeable à la translation et à la rotation, et change légèrement en cas de changements de l'angle de prise d'image ou d'échelle. Le système de couleur le plus subordonné est le RGB. D'autres systèmes tels que le HSV, le L *a*b*, ou le L *u*v* sont aussi couramment utilisés. La couleur est généralement décrite par un histogramme calculé dans un des divers espaces de couleur existant. La comparaison des images par les histogrammes est effectuée en utilisant une mesure de similarité comme l'intersection d'histogramme ou la distance Euclidienne. L'inconvénient majeur de l'histogramme est qu'il ne contient pas d'informations spatiales. Ainsi, pour pallier à cet inconvénient, plusieurs méthodes ont été proposées comme le découpage de l'image en zones d'intérêt, ou l'étude de la corrélation spatiale des couleurs (les correlogrammes) [7].

Les avantages de l'utilisation des histogrammes de couleur sont :

- ✓ L'extraction de l'histogramme est facile et rapide.
- ✓ Il est invariant à plusieurs transformations comme la translation, la rotation, le changement d'échelle et le point de vue de l'image.
- ✓ Généralement, il représente bien le contenu de l'image.

b) La texture

La texture est une information de plus en plus utilisée en indexation d'images et de la vidéo. Elle permet de combler un vide que la couleur est incapable de faire, notamment lorsque les distributions de couleur sont très proches, cette caractéristique permet de prendre en considération l'information spatiale.

Généralement il ya quatre types d'approches d'analyse de la texture: les approches statistiques, géométriques, spectrales et par modélisation [7].

L'inconvénient des méthodes qui exploite la texture est que l'extraction de cette caractéristique prendre beaucoup de temps.

c) La forme

La forme est utilisée pour caractériser les objets à l'intérieur de l'image. Généralement, la forme est décrite par des caractéristiques globales comme la taille (le périmètre et la superficie), l'excentricité et les moments ou par des caractéristiques plus précises comme les coins, les points de contours. Pour décrire la forme, Hu [8] a proposé un ensemble de sept moments invariants aux translations, aux rotations et aux changements d'échelle. Une amélioration de certaines caractéristiques invariantes aux transformations linéaires a ensuite été proposée par Reiss [9].

d) Le mouvement

Le mouvement est une caractéristique spécifique à la vidéo. Il est utilisé pour caractériser les déplacements des objets à l'intérieur de la séquence d'images qui constituent le plan, et pour caractériser les mouvements de la caméra. À partir des images qui constituent un plan, le mouvement de la caméra peut être estimé, ensuite le mouvement des objets peut être déterminé, parmi les méthodes les plus utilisées pour caractériser le mouvement, nous trouvons la technique basée sur la différence entre deux images consécutives [7].

1.2.5.2. Caractéristiques de haut niveau

Également connues sous le nom de caractéristiques logiques, dérivées ou sémantiques, les caractéristiques de haut niveau impliquent des degrés divers de sémantique représentés dans les images, la vidéo, et l'audio. Nous pouvons distinguer deux types de

a) Les caractéristiques objectives

Elles concernent l'identification des objets dans les images et l'action dans la vidéo. Un exemple de requête est « Trouvez une séquence vidéo contenant une baleine ». Pour répondre à des requêtes à ce niveau, le processus de recherche exige normalement une connaissance antérieure des objets [7].

b) Les caractéristiques subjectives

Elles sont des caractéristiques abstraites. Elles décrivent la signification et le but des objets ou des scènes. Nous pouvons subdiviser les caractéristiques en événements (par exemple, le jour de l'indépendance), en types d'activité (par exemple, le dessin), la signification émotive (par exemple, un sourire), le religieux (par exemple, une prière). L'interprétation complexe et le jugement subjectif peuvent être demandés à un expert dans le domaine d'application pour établir le lien entre le contenu de la vidéo et les concepts abstraits [7].

1.2.6. Vidéo et compression

MPEG (Moving Pictures Experts Group)

Selon [1], C'est un outil de compression qui a été crée en 1988, il est basé sur les similitudes existant entre plusieurs images successives.

MPEG-1 : destiné aux applications multimédia. Il permet la production des images de qualité équivalente au VHS tout en parvenant à descendre à un débit binaire de l'ordre de 1.2 Mbits/seconde (1.5 Mbits/seconde en incluant le son).

MPEG-2 : extension de MPEG-1 permettant d'obtenir une qualité d'image supérieure. Le but du MPEG-1 était Le MPEG-2 fut conçu pour traiter des séquences d'images entrelacées. Le but était de produire des images de la qualité d'un système vidéo composite avec un débit binaire de l'ordre de 4 à 8 Mbits/seconde ou des images de haute qualité avec un débit de 10 à 15 Mbits/seconde. Les domaines d'application principaux de MPEG-2 sont liés à la distribution de programmes vidéo : diffusion par satellite, télédistribution, Digital Vidéo Disc.

MPEG-3 : destiné à la télévision haute définition. Cependant, MPEG-2 s'est révélé tellement performant qu'il a rendu inutile le développement de MPEG-3.

MPEG-4 : destiné aux communications mobiles.

MPEG-7 : est un standard de description du contenu audio et vidéo [10] ,[11]. Le MPEG7 a proposé un certains nombres de caractéristiques standards pour décrire le contenu de la vidéo, bas niveau et haut niveau. Il standardise un ensemble de descripteurs D des entités de la production de l'audiovisuel, et des schémas de description qui décrivent les informations des différentes étapes d'élaboration de chaque entité de la production [12].

1.2.7. Vecteur descripteur et mesure de similarité

Selon [7], [13] et [14], Pour mesurer la similitude des images de deux séquences vidéo, l'approche typique, selon les auteurs, est de représenter chaque image par un vecteur de descripteur multidimensionnel qui est un vecteur caractéristique de l'image construite à partir des attributs extraits de l'image tels que la couleur, la texture, la forme et le mouvement. Il se présente généralement sous forme d'un vecteur $\mathbf{h}ij$ à la ***j***-ième des N composantes réelles du vecteur d'index $\mathbf{h}i$ de l'image i décrivant le contenu visuel et pouvant être de très grande dimension. Donc la similarité entre images est calculée par une fonction de distance appropriée dans un espace métrique multidimensionnel appliquée sur les vecteurs correspondants.

Mathématiquement, cette distance normalisée d est définie comme une fonction de distance entre les vecteurs $\mathbf{h}1$ et $\mathbf{h}2$. Cette distance d ($\mathbf{h}1$, $\mathbf{h}2$) vérifie les propriétés suivantes :

$$d(\mathbf{h}_1, \mathbf{h}_2) \geq 0$$
$$d(\mathbf{h}_1, \mathbf{h}_1) = 0$$
$$d(\mathbf{h}_1, \mathbf{h}_2) = d(\mathbf{h}_2, \mathbf{h}_1)$$
$$d(\mathbf{h}_1, \mathbf{h}_3) \leq d(\mathbf{h}_1, \mathbf{h}_2) + d(\mathbf{h}_2, \mathbf{h}_3)$$

Qui traduisent respectivement les propriétés de positivité, d'identité, de symétrie et d'inégalité triangulaire. Si ces propriétés ne sont pas – ou pas toutes – respectées, on parle plutôt de similitude entre vecteurs, avec pour notation s ($\mathbf{h}1$, $\mathbf{h}2$).

Les distances sont nombreuses dans la littérature, définies pour des valeurs scalaires, ensemblistes, vectorielles, etc. (comme différence absolue, cosinus, Harman, Jacquard, degré d'inclusion, Minkowski, Manhattan, Euclidienne, Hausdorff…). Nous présentons dans la suite les distances les plus couramment utilisées.

1.2.7.1. Distance entre vecteurs

a) La distance de Minkowski

Lorsque les données sont assimilées à des vecteurs, ce qui est souvent le cas, la distance de Minkowski est fréquemment employée. Elle est donnée par [14] :

$$d_{L_p}(\mathbf{h}_1, \mathbf{h}_2) = \sqrt[p]{\sum_{j=1}^{n} \left(h_{1,j} - h_{2,j}\right)^p}$$

Où p est un réel positif représente l'ordre.

Selon [14], en faisant varier *p* on obtient différentes types de fonctions comme :

b) La distance de Manhattan

La distance de Minkowski du premier ordre (*p=1*) est une distance de Manhattan :

$$d_{L_1}(\mathbf{h}_1, \mathbf{h}_2) = \sum_{j=1}^{n} |h_{1,j} - h_{2,j}|$$

c) La distance Euclidienne

La distance de Minkowski du deuxième ordre (*p=2*) est une distance Euclidienne :

$$d_{L_2}(\mathbf{h}_1, \mathbf{h}_2) = \sqrt[2]{\sum_{j=1}^{n} (h_{1,j} - h_{2,j})^2}$$

d) La distance de Tchebychev

Et quand *p* tend vers l'infini la distance de Minkowski tend vers la distance de Tchebychev:

$$d_{L_\infty}(\mathbf{h}_1, \mathbf{h}_2) = \max_{1 \le j \le n} |h_{1,j} - h_{2,j}|$$

1.2.7.2. Distance ou similitude entre histogrammes

Plus les distances ci-dessous, on peut mesurer la distance entre deux histogrammes par la distance *Euclidienne* et la distance *Manhattan.*

a) Similitude Swain

En recherche d'images par l'exemple, la première similitude entre histogrammes qui ait été utilisée est définie par [15]:

$$s_{Swain}(\mathbf{h}_1, \mathbf{h}_2) = \frac{\sum_{j=1}^{n} \min(h_{1,j}, h_{2,j})}{\sum_{j=1}^{n} h_{1,j}}$$

Elle a pour nom *intersection d'histogramme*. Dans cette expression, **h**1 est l'histogramme de l'image de premier plan, **h**2 l'image de deuxième plan. Il ne s'agit pas d'une distance, puisqu'elle ne respecte pas la propriété de symétrie. Pour y remédier, on peut utiliser l'expression suivante [16]:

b) La distance Smith

$$d_{Smith}(\mathbf{h}_1,\mathbf{h}_2)=\frac{\sum_{j=1}^{n}\min(h_{1,j},h_{2,j})}{\min\left(\sum_{j=1}^{n}h_{1,j},\sum_{j=1}^{n}h_{2,j}\right)}$$

c) La distance Kullbak-Leibler

Issue de la théorie de l'information, la divergence de Kullback-Leibler [17] permet de mesurer la dissimilarité basée sur l'entropie mutuelle de deux distributions :

$$d_{Kullback}(\mathbf{h}_1,\mathbf{h}_2)=\sum_{j=1}^{n}h_{1,j}\log\frac{h_{1,j}}{h_{2,j}}$$

d) La distance Jeffrey

Cependant, la version de Jeffrey lui est préférée pour son respect de la symétrie et de l'inégalité triangulaire [18]:

$$d_{Jeffrey}(\mathbf{h}_1,\mathbf{h}_2)=\sum_{j=1}^{n}\left(h_{1,j}\log\frac{2h_{1,j}}{h_{1,j}+h_{2,j}}+h_{2,j}\log\frac{2h_{2,j}}{h_{1,j}+h_{2,j}}\right)$$

Pour limiter la sensibilité au bruit de ces distances ou similitudes, on peut remplacer les histogrammes par les histogrammes cumulés. Mais ceci nécessite au préalable un ordonnancement des couleurs [19].

e) L'intersection d'histogramme

La distance entre deux histogrammes de couleur h et g est calculée pour chaque bande de couleur comme suit :

$$D(h,g)=\frac{\sum_{A}\sum_{B}\sum_{C}\min(h(a,b,c),g(a,b,c))}{\min(|h|,|g|)}$$

où |h| et |g| sont les magnitudes des histogrammes, lesquelles sont égales au nombre d'échantillons, et *a, b* et c représentent les bandes de couleur [7].

f) La distance quadratique

La distance quadratique entre deux histogrammes de la couleur *h* et *g* est calculée pour chaque bande de couleur comme suit:

$$D(h,j) = \sqrt{(h-g)^T A^{-1}(h-g)}$$

où *h* et *g* sont considérés comme des vecteurs de dimension K, et A = $[a_{ij}]$ est une matrice de dimension *K* x *K*, avec a_{ij} une distance quelconque mesurant la similitude entre la classe *i* et la classe *j*.

$$a_{ij} = 1 - d_{ij} / \max(d_{ij})$$

où *dij* est la distance L_2 entre la couleur *i* et *j* dans l'espace de couleur [7].

g) La distance EMD (Earth mover distance)

La distance EMD entre deux histogrammes est le travail minimal pour rendre les deux histogrammes identiques, en transportant le contenu des colonnes qui diffèrent d'un histogramme à l'autre. La distance entre deux histogrammes de la couleur *h* et *g* est calculée pour chaque bande de couleur comme suit :

où

$$EMD(h,g) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

- *n* est le nombre de cases de l'histogramme *h;*
- *m* est le nombre de cases de l'histogramme *g;*
- *dij = d(hi, gj)* est la distance entre *hi et gj,* C.-à-d. la distance entre la case *i de* l'histogramme *h* et la case *j* de l'histogramme *g;*
- *fij* est la quantité de masse transportée de la case *i* de l'histogramme *h* vers la case *j* de l'histogramme *g,* ou vice versa[7].

1.3. Segmentation de la vidéo

Dans le but de faciliter la représentation de contenus des documents vidéo dans la base d'index afin d'accélérer l'opération de la recherche de ces documents, la segmentation de la vidéo représente une étape primordiale qui permet d'atteindre ce but. Généralement il y'a deux types de segmentation :

1.3.1. Segmentation temporelle

D'après [20], La segmentation temporelle permet de faciliter l'identification et l'annotation des segments (temporels) ayant une unité sémantique. Ces unités sont obtenues d'un découpage temporel. En effet, des besoins d'informations varies tels que exprimés par les requêtes sur un concept X : « rechercher les segments vidéos montrant une image de X » et « rechercher les segments vidéos dans lesquels on parle de X », sont susceptibles de produire comme réponses deux unités sémantiques tout à fait différentes selon le media : image (l'unité peut être un plan par exemple) ou audio (l'unité peut être un segment audio).

La segmentation dite ***en plans***, il s'agit de découper une vidéo pour avoir une séquence d'images individuelles. Chaque plan est identifié par une image clé [21] et contient un ensemble d'images similaires. L'ensemble de ces images (Figure 6) forme ce que l'on appelle le ''résumé vidéo'' [22].

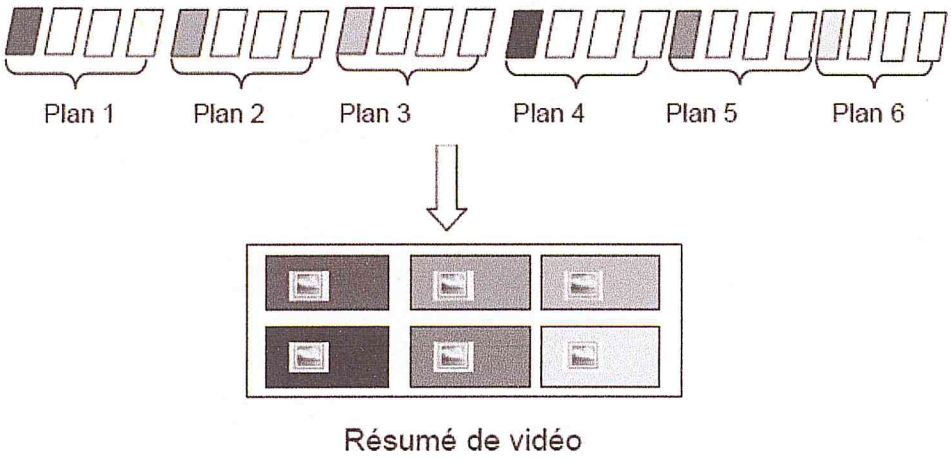


Figure 6 : Les plans, les images clefs et le résumé vidéo [22]

À partir d'un découpage ***en plans***, il ya deux types de segmentations supplémentaires :

- ✓ Le premier type est la ***micro-segmentation***. Elle est considérée comme une analyse plus détaillée du contenu d'un plan. Elle représente le contenu de la vidéo selon les mouvements de la caméra par des caractéristiques du contenu (apparition / disparition d'un objet ou d'une personne) [20].

- ✓ Le deuxième type permet d'extraire les images clés de chaque plan. Cette extraction est considérée comme une première étape pour la segmentation spatiale de l'image (découpage de l'image en régions, identification des objets) pour décrire son contenu [20].

1.3.2. Segmentation Spatiale

La segmentation spatiale consiste à partitionner le contenu de l'image en zones homogènes (couleur, texture, forme) et/ou correspondant à des objets (ou classes) La figure 7 récapitule les deux types de segmentations (temporelle et spatiale).

La segmentation spatiale permet de décrire le contenu visuel de la vidéo. Elle détermine les positions des objets visuels figure 8.

Il est possible d'utiliser les deux formes à la fois de segmentation (spatiale et temporelle), ce qui nous aide à assurer la continuité temporelle car les informations spatiales et temporelles sont le plus souvent employées dans un seul sens on exploite généralement l'information temporelle en premier lieu pour inférer ensuite une description spatiale [20].

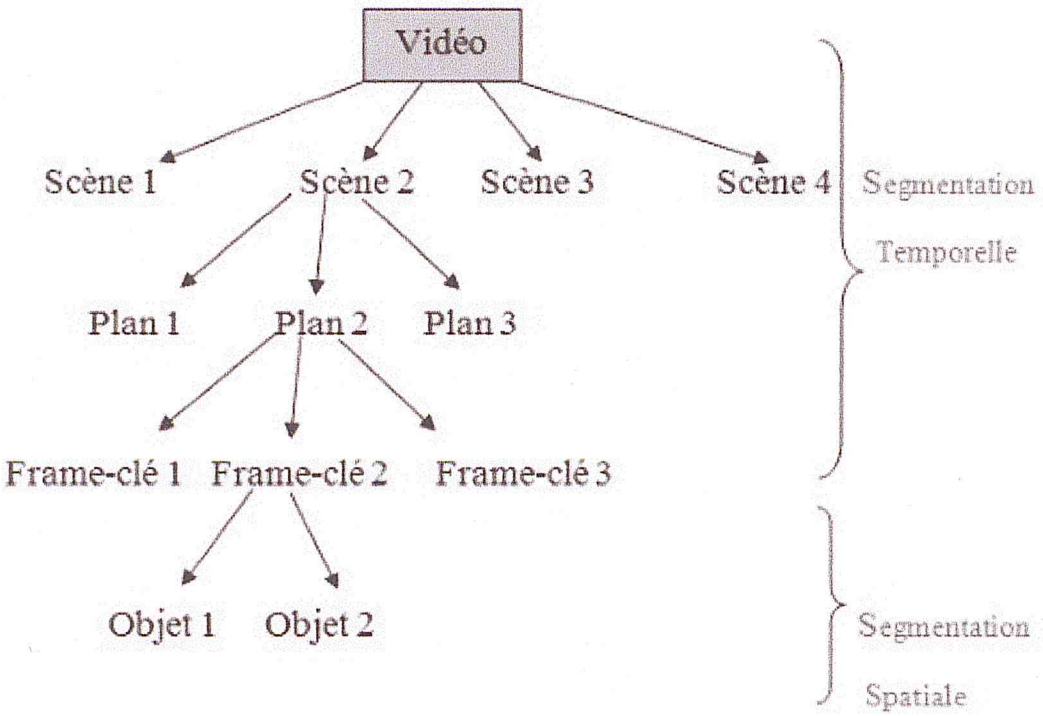


Figure 7: Exemple de structure spatiale et temporelle d'une vidéo [20]

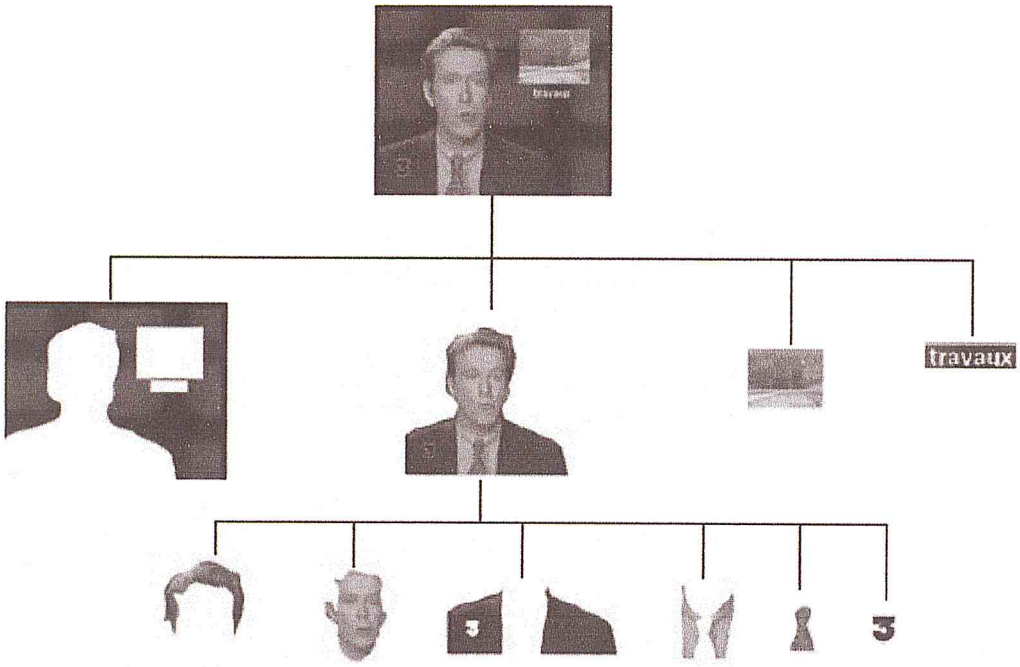


Figure 8 : Découpage spatial : segmentation en objets [20]

1.4. Panorama des méthodes de détection des changements de plans

Selon [5], la plupart des méthodes proposées pour résoudre le problème de la détection des changements de plans fonctionnent en deux étapes : le calcul d'une mesure de similarité entre deux trames successives d'une séquence vidéo, puis la comparaison de la valeur obtenue avec un seuil, afin de déterminer ou non la présence d'un changement de plans. Suivant ce principe, la détection d'un changement de plans est effective si la condition suivante est respectée :

$$D(I_t, I_{t-1}) > S$$

I_t représente l'image de la séquence vidéo obtenue à l'instant t, d une distance, et S un seuil.

Il y a plusieurs méthodes de détection de changement de plan qui se basent sur les séquences vidéo non compressées et compressées. Nous examinons les plus connues dans cette partie.

1.4.1. Séquences vidéo non-compressées

Nous proposons ici une classification des différentes méthodes selon le type ou la forme de l'information utilisée, les méthodes peuvent être basées sur les pixels, les histogrammes, un découpage en blocs et une information liée au mouvement.

1.4.1.1. Méthodes basées sur les pixels

Selon [7] et [23], La manière la plus simple de mesurer la différence entre deux images est de comparer les intensités ou les couleurs des pixels entre les deux images [24]. Selon cette méthode, il y a une coupure entre deux images qui se succèdent dans une vidéo, si la somme absolue des différences entre les pixels de l'image à l'instant t et ceux de l'image à l'instant $t+1$ est supérieur à un seuil fixé T.

Pour les images en niveaux de gris,

$$D(i, i+1) = \frac{\sum_{x=1}^{X} \sum_{y=1}^{Y} |P_i(x, y) - P_{i+1}(x, y)|}{XY}$$

Pour les images couleur,

$$D(i, i+1) = \frac{\sum_{x=1}^{X} \sum_{y=1}^{Y} \sum_{c} |P_i(x, y, c) - P_{i+1}(x, y, c)|}{XY}$$

où i et i + 1 sont deux images successives avec la dimension X × Y, *Pi* (x, y) est l'intensité valeur du pixel au point de coordonnées (x, y) dans l'image i,

c est l'indice pour les composantes de couleur (par exemple $c \in \{R, G, B\}$ dans le cas de système de couleurs RGB) et *Pi* (x, y, c) est la composante de couleur du pixel en (x, y) dans l'image i.

L'inconvénient de cette méthode est sa sensibilité aux mouvements des objets et de la caméra, car en utilisant le changement de la moyenne, il est impossible de faire la différence entre un grand changement dans une petite région de l'image et un petit changement dans une grande région. Zhang et al. [25] ont proposé une amélioration qui consiste à déterminer le pourcentage des pixels qui ont changé considérablement entre deux images et comparé à un seuil T_2. Si le pourcentage de changement de pixels DP (i, i + 1, x, y) est supérieur à T_2, une coupure est détectée.

$$\mathrm{DP}(i, i+1, x, y) = \begin{cases} 1 & |P_i(x, y) - P_{i+1}(x, y)| > T_1, \\ 0, & \end{cases}$$

$$D(i, i+1) = \frac{\sum_{x=1}^{X} \sum_{y=1}^{Y} \mathrm{DP}(i, i+1, x, y)}{XY}.$$

Leur méthode utilise un filtre moyen 3x3 pour réduire le bruit et l'effet du mouvement de la caméra. Bien qu'elle apporte une amélioration, cette méthode est toujours sensible au mouvement des objets et de la caméra.

1.4.1.2. Méthodes basées sur les histogrammes

D'après [7] et [23], Quelques méthodes ont été proposées pour pallier au problème du mouvement de la caméra et des objets. Ces méthodes comparent les caractéristiques globales de chaque image au lieu de comparer chaque pixel individuellement. Nagasaka et Tanaka [26] ont proposé l'utilisation de l'histogramme à niveau de gris qui se fait en calculant une distance entre les histogrammes des images basée sur l'équation suivante [27]:

$$\left(\sum_{v=0}^{v} |H(I_t, v) - H(I_{t-1}, v)|\right) > T$$

Toutefois, la méthode n'était pas robuste en présence de bruit momentané, comme le flashe d'un appareil photo ou le mouvement d'un grand objet. Nagasaka et Tanaka [26] ont également proposé une méthode basée sur la comparaison de l'histogramme de la couleur définit dans l'équation suivante *H64(I,v)* :

$$\left(\sum_{v=0}^{64} |H_{64}(I_t, v) - H_{64}(I_{t-1}, v)|\right) > \mathrm{T}$$

Ils ont proposé d'utiliser un code couleurs de 6 bits obtenus en prenant les deux bits les plus significatifs de chaque composante RGB ce qui donne un code à 64 couleurs. Ils utilisent la loi de Chi deux *X2* pour mesurer la différence entre deux distributions liées.

Selon Gargi et al. [28], Nagasaka et Tanaka [26] et Lienhart [29], une simple comparaison entre les histogrammes de la couleur (RGB ou YUV), avec chaque bande quantifiée à 2^b valeurs différentes, est une méthode efficace pour détecter les frontières des plans.

Le principal défaut de ce type de méthodes est leur caractère purement global.

1.4.1.3. Méthodes basées sur les blocs

Selon [7] et [23], La détection des changements de plans dans ce procédé peut aussi être effectuée en utilisant les *pixels* et les *histogrammes*.

Chaque image i est divisée en blocs b qui sont comparées avec leurs blocs correspondants dans i + 1. Typiquement, la différence entre i et i + 1 est mesurée par

$$D(i, i+1) = \sum_{k=1}^{b} c_k \, \mathrm{DP}(i, i+1, k),$$

où c_k est un coefficient prédéterminée pour le bloc k et DP (i, i + 1, k) est une valeur de correspondance partielle entre les blocs de k dans i et i + 1 images.

Dans [30] les blocs correspondants sont comparés en utilisant un rapport de vraisemblance de Yakimovsky

$$\lambda_k = \frac{\left[\frac{\sigma_{k,i} + \sigma_{k,i+1}}{2} + \left(\frac{\mu_{k,i+1} - \mu_{k,i+1}}{2}\right)^2\right]^2}{\sigma_{k,i} \cdot \sigma_{k,i+1}},$$

où $\sigma_{k,i} + \sigma_{k,i+1}$ sont les valeurs moyennes pour les deux blocs correspondants d'intensité k dans les images consécutives i et i + 1, et $\mu_{k,i+1} - \mu_{k,i+1}$ sont leurs écarts, respectivement.

Ensuite, le nombre de blocs pour lesquels le rapport de vraisemblance est supérieur à un seuil T_1 est compté,

$$\mathrm{DP}(i, i+1, k) = \begin{cases} 1 & : \lambda_k > T_1, \\ 0 & , \end{cases}$$

Une coupe est déclarée lorsque le nombre de blocs modifiés est assez grand, c.-à-d (i, i + 1) est supérieur à un seuil T_2 donné et $c_k = 1$ pour tout k.

Les approches à base d'histogrammes consistent à appliquer un traitement d'image par blocs. Alors on calcul pour chaque bloc son histogramme et on fait la comparaison avec celui du bloc correspondant dans l'image précédente. La formule suivante donne un exemple qui montre le calcule de la similarité [31]:

$$\sum_{b=1}^{B} \sum_{n=1}^{N} \sum_{c=1}^{C} \left| H(I_t^b, n, c) - H(I_{t-1}^b, n, c) \right| > S$$

Tel que :

H : la valeur de l'histogramme. *It :* l'intensité de l'image en cours.

n : le nombre d'images utilisées. *c* : le canal sélectionné.

S : le seuil utilisé. *b* : le bloc sélectionné.

Ce type de méthode est un bon compromis entre les approches purement locales et les approches globales.

1.4.1.4. Méthodes basées sur le mouvement

Cette méthode se caractérise par son invariance aux changements dans l'illumination globale du contenu de l'image. Fernando et al. [32] ont exploité le fait que les vecteurs de mouvement sont de nature aléatoire pendant une coupure de plan. La méthode calcule le vecteur de mouvement moyen entre deux images et la distance Euclidienne par rapport au vecteur moyen pour tous les vecteurs de mouvement. Ils déduisent qu'il y a une coupure s'il y a une grande augmentation dans la distance Euclidienne.

1.4.2. Séquences vidéo compressées

Ces séquences vidéo compressées supposent généralement une compression vidéo suivant les normes MJPEG ou MPEG, nous décrivons la méthode la plus utilisé baser sur les coefficients de la DCT (*Discrete Cosine Transform* ou en français Transformée en Cosinus Discrète).

1.4.2.1. Méthodes basées sur les coefficients de la DCT

Ces approches sont utilisées généralement pour une séquence vidéo compressée car les informations dans ces séquences vidéo sont codées par les coefficients DC et AC issus de la DCT [33].

Pour une image à deux dimensions la DCT est définie par l'équation suivante :

$$C(u,v) = a(u)a(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)cos\left(\frac{\pi(2x+1)u}{2N}\right)cos\left(\frac{\pi(2y+1)v}{2N}\right)$$

$$a(u)= \sqrt{\frac{1}{N}} \quad \text{si } u=0 \quad \text{ET} \quad a(u)= \sqrt{\frac{2}{N}} \quad \text{si } u\neq 0$$

Tel que :

N : la taille de la séquence.

f(x, y) : la valeur de l'élément aux indices x et y dans la séquence de sortie.

C(u,v) : la valeur de l'élément aux indices u,v dans la séquence de sortie

Donc le calcul du premier coefficient nous donne le résultat suivant :

$$C(u,v) = a(0)a(0)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)cos\left(\frac{0}{2N}\right)cos\left(\frac{0}{2N}\right) = \frac{1}{N^2}\sum_{y=0}^{N-1} f(x,y)$$

On peut voir que le C(0,0) est la valeur moyenne de l'image, donc elle représente l'énergie moyenne, on l'appel alors : le coefficient DC. Les autres coefficients on les appels les coefficients AC [30].

Donc l'idée ici est de détecter les changements de plans par analyse de ces coefficients dans les trames successives [33]. Le principe est de calculer le produit scalaire entre deux vecteurs de caractéristiques qui comportent les coefficients DC. La détection de changement de plan est confirmée par le produit scalaire (s'il est faible) et l'analyse des histogrammes couleur, ou bien une comparaison bloc-à-bloc est faite entre les deux images successives, toujours via l'analyse des coefficients de la DCT.
Un changement de plans est effectué s'il y a un nombre de paires de blocs suffisamment différents [34].

On peut utiliser aussi, la moyenne et la variance de l'intensité des images comme des mesures pour détecter le changement de plans, elles sont calculées à partir des coefficients DC, ces mesures peuvent être calculées sur l'image globale ou alors selon les directions horizontale et verticale.

1.5. Conclusion

La détection des changements de plans est un des traitements des vidéos nécessaire pour caractériser une séquence vidéo.

Nous avons présenté dans ce chapitre, les notions de base sur la vidéo et un certain nombre de méthodes de segmentation en plans de vidéo; ceci nous a donné une vue globale sur les techniques de détection de changement de plans pour les vidéos compressées et non compressées. Le chapitre suivant sera consacré à l'étude détaillée des travaux relatifs aux méthodes de segmentation temporelle basées respectivement sur des histogrammes couleur et sur les informations de mouvement.

Chapitre 2 : Les approches Retenus

2.1. Introduction

Suite aux notions de base présentées précédemment, ce chapitre s'intéresse à décrire les différentes étapes de deux techniques choisi de segmentation en plans de vidéo basée sur la méthode d'histogramme de couleur et la méthode du mouvement. Ces deux méthodes sont bien représentés le contenu de l'image et elles sont bien invariant à translation, rotation et changement d'échelle. Nous présentons en premier lieu les descripteurs visuels utilisés. Par la suite, nous décrivons l'algorithme de chaque méthode utilisée dans notre application.

2.2. Les descripteurs visuels utilisés

Pour représenter le contenu des images, des caractéristiques de bas niveau sont extraites sur chacune des images sont celles de la couleur par l'histogramme de couleur et la caractéristique du mouvement qui est spécifique à la vidéo. Nous allons présenter brièvement ces descripteurs les plus utilisées en recherche de la vidéo.

2.2.1. La couleur

La couleur est cet aspect de lumière visible, par lequel un être humain distingue entre différentes répartitions spectrales de l'énergie de lumière. En utilisant son système de vision, l'être humain interprète les couleurs grâce aux quantités de lumière de longueurs d'onde variées que les objets autour de lui émettent ou réfléchissent. En fait, la plupart des couleurs sont dues non à des mélanges de longueurs d'onde, mais à des soustractions. La lumière blanche du soleil étant partiellement absorbée par des pigments qui absorbent certaines longueurs d'onde et ne laissent passer que leur complément, ce qui produit la sensation de couleur. L'être humain peut différencier jusqu'à environ deux millions de couleurs différentes. La figure 9 montre le spectre visible pour l'œil humain.

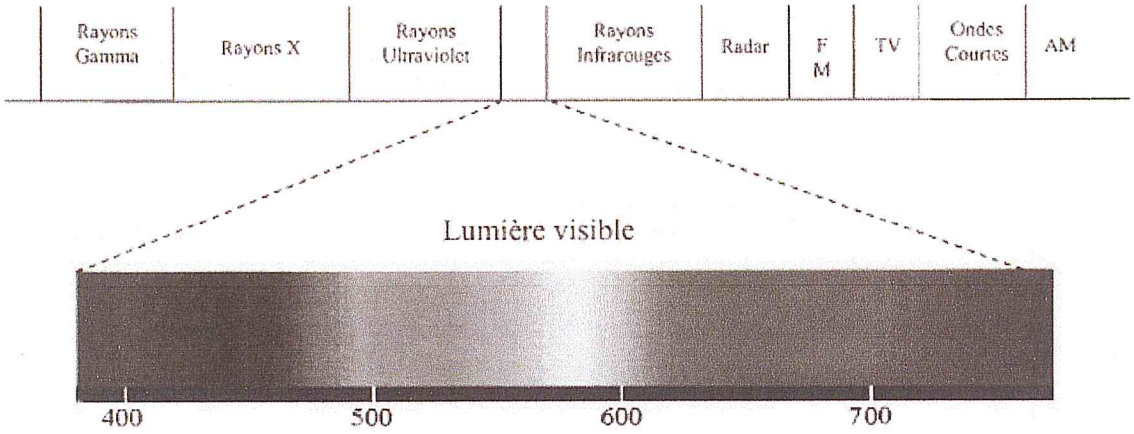


Figure 9: Le spectre visible [7].

De nos jours, presque la totalité des images et des vidéos présente dans les différents médias sont en couleur, car elles sont plus réalistes et satisfaisantes à l'œil humain. La nécessité de produire, de stocker et de transmettre des documents colorés (image ou vidéo) a conduit à imaginer des systèmes cohérents pour représenter fidèlement les couleurs qui les composent. Ces systèmes sont appelés les espaces de couleurs. Ils sont appelés ainsi, car la variation des différentes couleurs peut être représentée dans un espace tridimensionnel, où chaque point (dans cet espace) représente une couleur différente. Ce nuage de points de couleur différente constitue un espace de couleur. En conséquence, l'espace de couleur est une notation par laquelle nous pouvons spécifier les couleurs, c'est-à-dire la perception humaine du spectre électromagnétique visible.

Plusieurs espaces de couleurs ont été employés pour la représentation de couleur basée sur les concepts perceptuels. Nous pouvons citer les espaces RGB, CMY, HSV, ClELab, etc. Il n'y a aucun accord sur le meilleur espace de couleur. Cependant, ses caractéristiques désirables sont la perfection, l'uniformité, la compacité, et il doit être orienté utilisateur.

- ✓ La perfection signifie qu'il doit inclure toutes les couleurs différentes perceptibles.
- ✓ L'uniformité signifie que la proximité mesurée parmi les couleurs doit être directement rapprochée de la similitude perceptuelle ou psychologique entre ces couleurs.
- ✓ La compacité signifie que chaque couleur présente une différence perceptrice des autres couleurs.

Dans ce qui suit, nous allons donner plus de détails sur les espaces de couleur RGB [7].

2.2.1.1. L'espace RGB *(Red, Green, Blue)*

C'est l'espace de couleur de base. Il est très utilisé dans les systèmes de télévision et les applications informatiques. La représentation des couleurs dans cet espace donne un cube appelé « cube de Maxwel », comme illustré dans la figure 10. Le système de couleur RGB est un système de couleur additif, c.-à-d. que les couleurs sont obtenues par le mélange des trois couleurs de base qui sont le rouge, le bleu et le vert. La représentation numérique la plus fréquente de cet espace de couleur est des valeurs allant de 0 à 255.

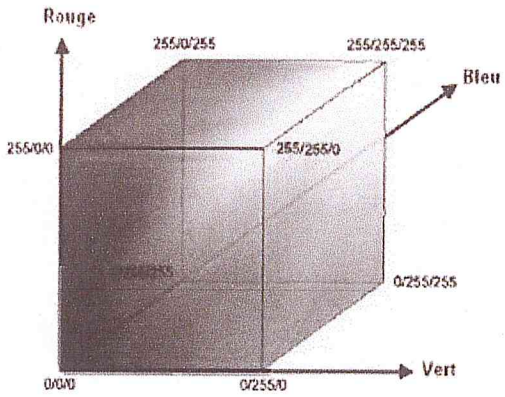


Figure 10: L'espace de couleur RGB [7].

L'avantage de l'espace de couleur RGB est qu'il est conceptuellement simple. En contrepartie, il a l'inconvénient d'être perceptuellement non uniforme, c.-à-d. il n'y a pas de corrélation entre la différence perçue entre deux couleurs différentes, et la distance Euclidienne qui sépare ces deux couleurs. De plus, l'espace de couleur RGB ne tient pas compte des particularités de la perception visuelle des couleurs, il n'est pas indépendant du matériel utilisé, et il n'est pas très intuitif pour les utilisateurs non initiés [7].

2.2.2. L'histogramme de la couleur

Selon [7], L'histogramme de la couleur est une représentation de la distribution des couleurs dans une image. Il est produit en découpant d'abord les bandes de l'espace de couleur utilisé dans un certain nombre de cases, puis en comptant le nombre de pixels dans chaque case. Formellement, l'histogramme de couleurs est défini comme suit:

$$h_{A,B,C}[a,b,c] = N.\ Prob\{A=a,\ B=b,\ C=c\}$$

où A, B et C représente les bandes de couleur dans l'espace de couleur choisie (RGB, HSV, etc.), et N est le nombre de pixels dans l'image. La figure 11 illustre un exemple d'une image

et ses différents histogrammes de la couleur, c.-à-d. un histogramme pour chaque bande de couleur.

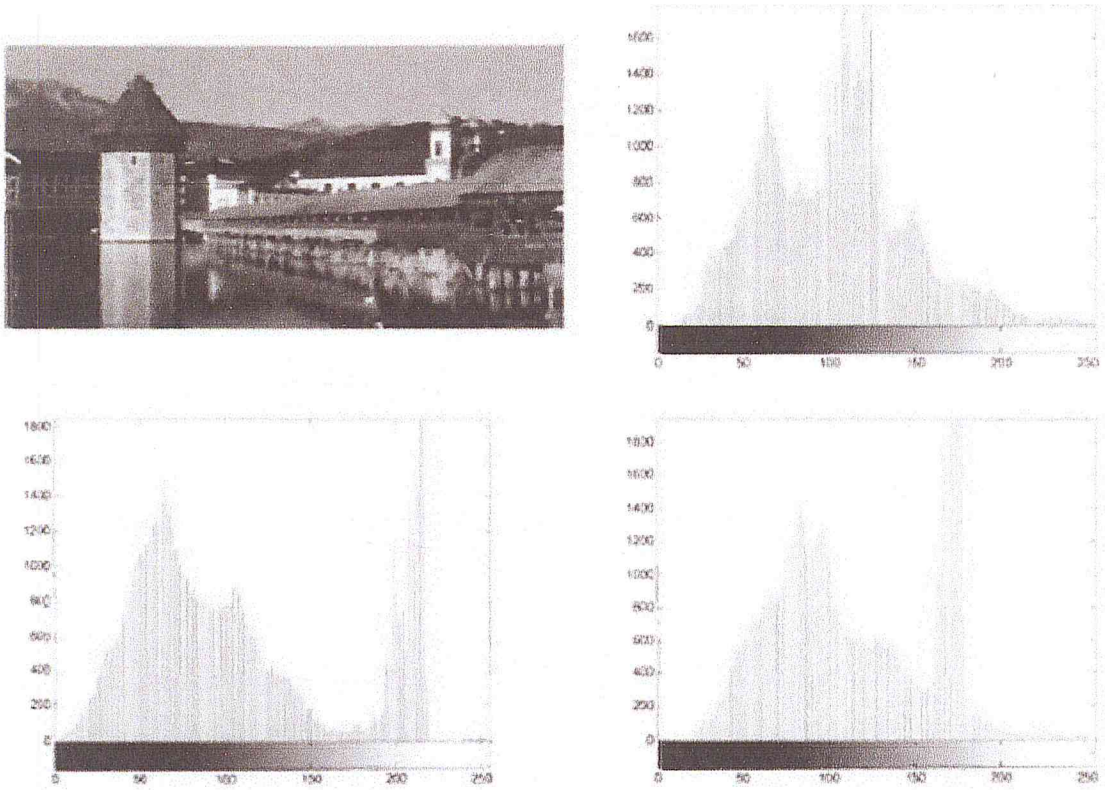


Figure 11: Les différents histogrammes d'une image couleur [7].

Les étapes d'extraction du descripteur histogramme de couleur sont représentées ci-dessous :

a) En utilisant l'espace de couleur RGB.

b) Formellement, l'histogramme de la couleur est défini comme suit:

$$h_{A.B.C}[a,b,c] = N.\ Prob\{A=a,\ B=b,\ C=c\}$$

où *A, B* et C représentent les bandes de couleur dans l'espace de couleur RGB, et *N* est le nombre de points dans l'image.

c) En décomposant l'espace de couleur en 27 sous-espaces,

d) En divisant les intensités dans chaque bande de couleur en trois parties égales.

e) Le résultat est un vecteur de 27 cases seulement.

La figure 12 montre une illustration du découpage de l'espace RGB.

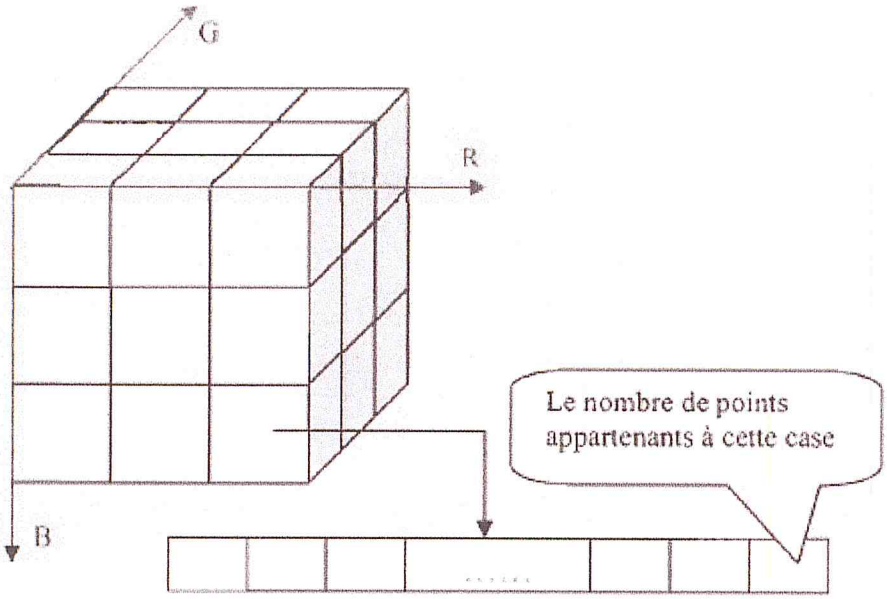


Figure 12 : L'histogramme de la couleur dans l'espace RGB [7].

L'histogramme de la couleur est couramment utilisé dans la recherche d'images et de la vidéo par le contenu en raison de ses nombreux avantages. Micheal Swain et Dana Ballard [35] sont parmi les premiers à l'avoir utilisé en recherche d'images, en 1991.

Parmi les avantages de son utilisation, nous pouvons citer les suivantes:

- L'extraction de l'histogramme est facile et rapide.
- Il est invariant à plusieurs transformations comme la translation, la rotation, le changement d'échelle et le point de vue de l'image.
- Généralement, il représente bien le contenu de l'image.
- Différentes mesures de similarité peuvent y être appliquées.

Par contre, afin que l'histogramme soit utilisé efficacement, il faut d'abord régler un certain nombre de questions.

• L'histogramme de la couleur ne contient pas d'informations spatiales. En d'autres termes, il ne donne aucune information sur l'emplacement des objets dans l'image. En effet, l'histogramme nous informe sur les couleurs présentes dans l'image et la proportion occupée par chacune. Cependant, il ne fournit aucune information sur la couleur d'une zone en particulier de l'image, ni sur l'endroit où une couleur est présente dans l'image, ni sur le fait qu'une couleur correspond à une seule région ou à des régions disjointes. La figure 13 illustre bien ce problème.

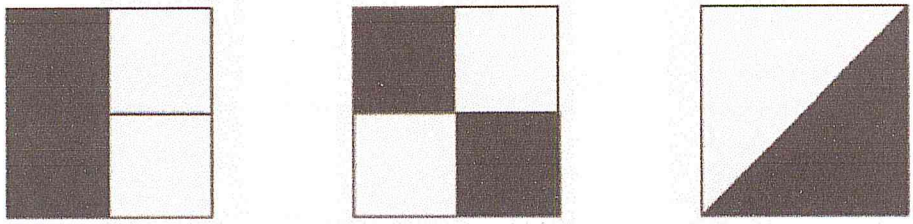


Figure 13 : Des images perceptuellement différentes avec des histogrammes de la couleur identique [7].

Parmi les solutions proposées pour résoudre ce problème, Hadjidemetriou et al. [36] ont utilisé l'histogramme de l'image ensemble, avec les différences entre les histogrammes de la même image à différentes résolutions pour encoder les informations spatiales. Cependant, l'efficacité de cette méthode dépend de la forme et de la texture de l'image.

• L'histogramme de la couleur n'est pas invariant à l'illumination. Dans l'espace de couleur RGB, la distribution des couleurs dans une image change proportionnellement avec l'illumination. Ainsi, l'histogramme de la même image change selon son degré d'illumination. Parmi les solutions proposées pour résoudre ce problème, il y a l'utilisation d'un pourcentage de couleur des pixels voisins [37] ou l'utilisation des moments invariants [38] ,[39].

• L'histogramme de la couleur a une grande dimension. Sans quantification, il est une caractéristique de dimension très élevée. Par exemple, si nous avons dans chaque bande de couleur 256 intensités différentes, nous obtenons un histogramme de 2563 cases, dont la plupart sont vides. Ceci constitue un handicap majeur: la recherche n'est plus précise, le temps de recherche devient inacceptable, et la mémoire nécessaire est énorme. Une solution évidente pour réduire la dimension de l'histogramme de la couleur est de réduire la gamme des couleurs. Cela peut être fait, parce que l'œil humain ne fait pas la différence entre des couleurs proches. Par exemple, nous pouvons remarquer qu'il est très difficile de distinguer la différence entre les couleurs de la première et de la deuxième case dans la Figure 14.

1	2	3	4	5	6
RGB = 255,0,0	RGB = 250,0,0	RGB = 212,0,0	RGB = 170,0,0	RGB = 127,0,0	RGB = 42,0,0

Figure 14 : La différence perceptuelle entre les couleurs [7].

Ainsi, au lieu de diviser chaque bande en 256 couleurs différentes, on peut la diviser en *n* intervalles (couleurs différentes) où *n* est beaucoup plus petit que 256 (ex. $n = 8$). À la fin, on se retrouve avec un histogramme de n^3, ce qui est beaucoup plus petit que 256^3. Il faut bien sûr trouver un découpage judicieux qui assure que chaque couleur tombe dans une case à part. Parmi les solutions pour résoudre ce problème de dimension, Wong et al. [40] ont proposé de réduire le nombre des couleurs utilisées à extraire l'histogramme de la couleur en produisant une palette commune. Ils ont produit cette palette en regroupant les couleurs perceptuellement semblables. Deng et al. [41] ont proposé une autre méthode basée sur l'observation qu'un petit nombre de couleurs est habituellement suffisant pour caractériser l'information de couleur dans une région de l'image. Ils ont groupé les couleurs dans une région donnée dans un certain nombre de couleurs représentatives. Puis, ils ont extrait le vecteur caractéristique à partir des couleurs représentatives et leur distribution dans les régions. Kherfi et al. [42] ont proposé la division de chaque bande de l'espace de couleur RGB en trois cases. Ainsi, ils ont réduit radicalement la dimension du vecteur caractéristique de l'histogramme de la couleur de 224 à 27 cases.

• L'histogramme de la couleur a l'inconvénient des similarités entre les cases. En d'autres termes, lors de la quantification d'une image, des pixels avec des couleurs qui sont perceptuellement très semblables peuvent être placés dans des cases d'histogramme différentes, mais voisines. Cela peut mener que, la différence entre deux histogrammes est beaucoup plus grande que la différence perceptuelle entre deux images.
Parmi les solutions proposées pour résoudre ce problème, c'est l'ajout à chaque case de l'histogramme les couleurs qui se situent à sa frontière. Ainsi, El-Feghi et al. [43] ont proposé une méthode basée sur la contribution de la couleur de chaque pixel dans l'image à toutes les cases de l'histogramme, à travers l'utilisation des fonctions d'ensemble flou.
Les mesures de similarité les plus utilisées avec les histogrammes sont la distance Euclidienne, la distance de Mahalanobis, l'intersection d'histogramme, la distance quadratique, la distance EMD (Earth mover distance), la distance Jeffrey, la distance Kullbak-Leibler et la distance Smith qui sont détaillés dans le chapitre précédent.

2.2.3. Le mouvement

La vidéo est constituée d'une séquence d'images. Décrire le contenu dynamique à l'intérieur de la vidéo est le résultat des changements qui se produisent à travers le temps à l'intérieur du contenu des images qui constituent la vidéo. Ces changements sont le résultat du mouvement de la caméra, du mouvement des objets à l'intérieur de la vidéo, ou d'une combinaison des deux. Les mouvements de la caméra les plus courants sont: la translation, la rotation et le zoom. La figure 15 présente une illustration de ces différents types de mouvement de la caméra [7].

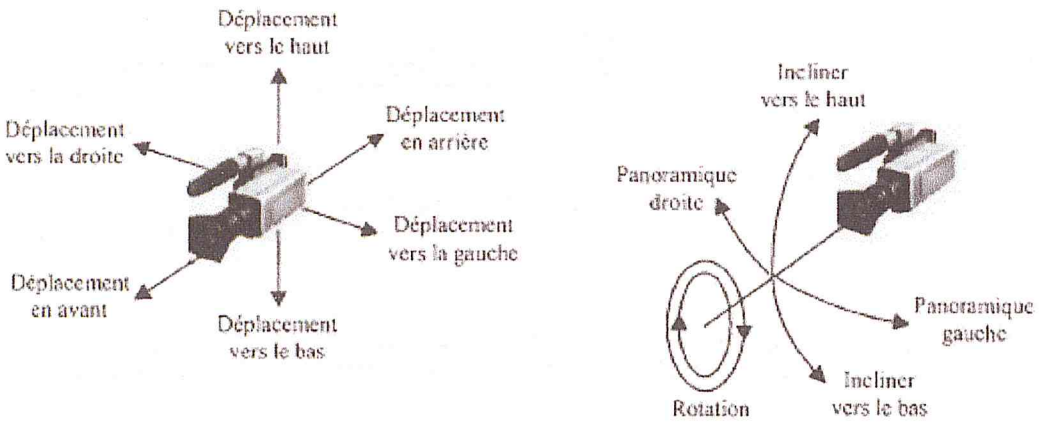


Figure 15 : Les mouvements courants de la caméra [7].

Nous allons présenter dans ce qui suit les différentes caractéristiques du mouvement et les méthodes utilisées pour les extraire.

2.2.3.1. Les différentes caractéristiques du mouvement

D'après [7], Les différentes caractéristiques du mouvement dans une vidéo sont:

a. L'activité du mouvement *(Motion activity)* : Cette caractéristique capture la notion d'intensité du mouvement d'une façon globale dans le plan, en utilisant les différents champs de vecteurs de mouvement extraits d'un plan [44] ,[45], [46] .Elle inclut les attributs suivants:

- L'intensité de l'activité: le niveau d'activité peut être haut ou bas selon le sujet de la vidéo. Par exemple, une vidéo d'une course automobile comporte un mouvement intense alors que le mouvement dans une vidéo de diner familial est peu intense.
- La direction de l'activité: elle indique la direction dominante de l'activité dans le plan.
- La distribution spatiale de l'activité: elle indique le nombre et la taille des régions actives dans les images qui constituent le plan.

• La distribution temporelle de l'activité: elle indique la variation de l'activité toute au long de la durée du plan.

b. Le mouvement de la caméra ***(Camera motion)*** : Cette caractéristique indique le type de mouvement de la caméra dans le plan (translation, zoom, rotation, etc.), son amplitude et sa localisation temporelle dans le plan. Les méthodes utilisées pour extraire cette caractéristique reposent sur des modélisations paramétriques 2D ou 3D des mouvements de la caméra [47], [48], [49] ;

c. Les paramètres de déformation ***(Warping parameters)*** : On estime les paramètres d'un modèle mathématique qui représente le panorama par une application. Cette application permet, à partir d'une seule image du panorama, de trouver les autres images de ce même panorama. Notons que le panorama est un mouvement de la caméra du type zoom, affine, etc. [50], [51], [52], [53];

d. La trajectoire du mouvement ***(Motion trajectory)*** : Elle décrit les déplacements des objets dans le temps. En général, c'est les positions successives dans le temps du centre de gravité d'un objet [49], [54] ;

e. Le mouvement paramétrique ***(Parametric motion)*** : Cette caractéristique permet d'extraire les objets ayant des mouvements similaires et qui subissent des rotations ou des déformations. Elle utilise le même modèle paramétrique de mouvement que les paramètres de déformation [50], [51], [52], [53];

Nous nous intéressons dans notre travaille par la caractéristique de l'activité du mouvement qui quantifie l'intensité du mouvement dans les vidéos.

2.3. Segmentation de la vidéo par les histogrammes de couleurs

Nous présentons la méthode qui sera utilisée pour réaliser la segmentation d'une vidéo par la méthode d'histogramme de couleur.

2.3.1. La transformation réversible de couleur

Un codage typique des images consiste à attribuer à chaque pixel un triplet unique associé aux composants rouge (Red), vert (Green) et bleue (Blue) – RGB. La transformation en question associe chaque triplet (R, G, B) à un entier unique F(R, G, B).

La caractéristique principale de cette transformation est qu'elle est bijective, c'est-à-dire qu'elle permet de retrouver la couleur (R, G, B) à partir de F(R, G, B) [55].

Soit *n* un entier codé sur 8 bits, et *mi* sa représentation au $i^{ème}$ bit. On définit le vecteur *U(n)* par :

$$U(n) = \sum_{i=1}^{8} m_i 2^{3(i-1)}$$

La fonction réversible de couleur est définit comme suit :

$$F(R, G, B) = 4U(G) + 2U(R) + U(B)\ldots\ldots (*)$$

R, G, B étant codés sur 8 bits. La fonction *F* est bijective de [0..255] vers $[0..2^{24}[$ [57]. Cette transformation est basée sur une décomposition binaire des composantes de couleurs, réalisée grâce au vecteur *U*. Remarquons que puisque *U* ne peut prendre en argument que les valeurs de 0 à 255, les *U(n)* peuvent être calculées une fois pour toute.

Le temps de calcul sera ainsi plus rapide.

On constate aussi que dans la relation (*), le composant vert est plus pesant que le rouge et le bleu. Ceci est dû au fait que l'œil humain est plus sensible aux variations de la couleur verte qu'aux rouge ou au bleu.

2.3.2. L'histogramme de couleur

Afin de générer un histogramme de couleur à partir de la transformation réversible, on réalise une quantification du *F(R, G, B)* de chaque pixel en *M* niveaux [57]. L'histogramme à *M* niveaux est établi par la relation :

$$H_F(k) = \sum \delta(Q_F(F(R_{i,j}, G_{i,j}, B_{i,j})) - k), \quad \text{pour } 0 \le k \le M$$

$$\delta(i-j) = \begin{cases} 1 & pour \quad i = j \\ 0 & pour \quad i \neq j \end{cases}$$

Où $Q_F(\,)$ représente la fonction de quantification qui quantifie $F(R_{i,j}, G_{i,j}, B_{i,j})$ en une valeur entre 0 à 255.

La différence entre deux histogrammes successives est un histogramme définit par la distance Euclidienne :

$$dH_i = \sqrt{\sum_{j=0}^{M-1} \left(H_{i-1}(j) - H_i(j)\right)^2} \qquad \text{pour } i=1,2,....L-1$$

Avec *L* le nombre total d'images dans la séquence vidéo.

Si la différence dH_i entre l'image courante et la précédente est supérieure à un seuil optimal *T (fixé de manière empirique)*, l'image courante est considérée comme un *cut*, tandis que la franche vidéo est générée automatiquement.

2.4. Segmentation de la vidéo par le mouvement

Dans cette partie nous allons citer la méthode la plus adaptées pour la détection de mouvement.

2.4.1. Détection basée sur la différence entre deux images consécutives

Comme son nom l'indique, elle consiste à soustraire une image acquise au temps t_n d'une autre au temps t_{n+k}, où k est habituellement égal à 1. Ainsi, l'image résultante sera vide si aucun mouvement ne s'est produit pendant l'intervalle de temps observé car l'intensité et la couleur des pixels seront presque identiques. Par contre, si un mouvement a eu lieu dans le champ de vue, les pixels frontières des objets en déplacement devraient changer de valeurs, révélant alors la présence d'activité dans la scène [2].

Nous calculons en premier lieu la différence pour chacun des pixels et pour chaque trame de l'image, ensuite nous calculons les étiquettes de mouvement en comparant le résultat t(x) de la différence avec un seuil T. Le choix du seuil de décision doit tenir compte du bruit et des changements de luminosité.

$$\text{Max}\ (|\ It(x, y) * c - I\ t\text{-}1\ (x, y) * c\ |, \quad c=(R, G, B)) >= T$$

a. Calcul de la différence

Algorithme de [59]

Pour chaque pixel (x,y) :

$0(x,y) \leftarrow 0$

Pour chaque trame t

Pour chaque pixel x :

$t(x,y) \leftarrow |I_t(x,y) - I_{t-1}(x,y)|$

b. Calcul des étiquettes de mouvement

On calcul les étiquettes de mouvement par comparaison entre la différence et le seuil T

Suite de l'algorithme de [59]

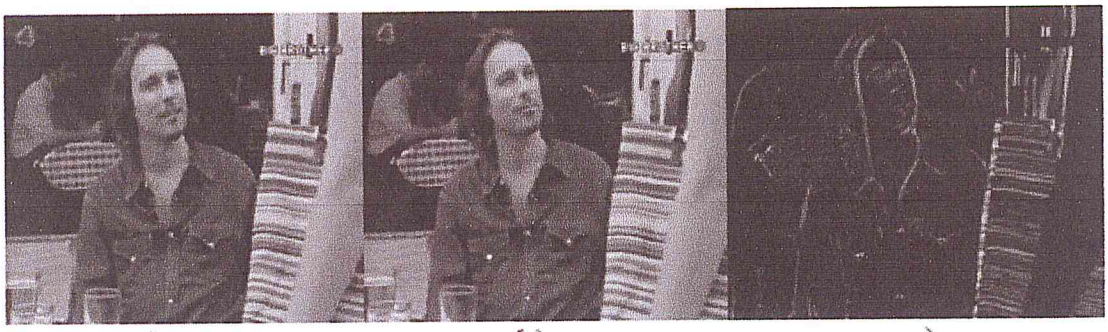
Pour chaque trame t

Pour chaque pixel x

Si $t(x,y) > T$

Alors E $t(x,y) \leftarrow 1$

Sinon E $t(x,y) \leftarrow 0$



a) b) c)

Figure 16 : Détection du mouvement par la méthode de soustraction d'images consécutives [2]

a) Image t0, b) Image t1, c) Détection de mouvement

2.5. Conclusion

Dans ce chapitre nous avons présenté la description générale de chaque descripteur utilisé dans les techniques que nous avons utilisées pour la segmentation en plans de la vidéo. Le présent système se base sur l'histogramme de couleur et le mouvement.

Les algorithmes présentés sont spécifié et génère les plans de la vidéo à partir d'un ensemble d'images obtenu par le découpage de vidéo. La similarité entre les images d'un plan vidéo est obtenue par le calcul de la distance de similarité.

Dans ce qui suit, nous montrons les résultats de la segmentation en plans de la vidéo obtenue pour différentes vidéos afin de comparer les deux méthodes.

Chapitre 3 : Expérimentation Et Discussion

3.1 Introduction

Ce chapitre comporte les algorithmes implémentés pour la réalisation des approches adoptées, l'environnement de travail, l'interface du système développé, les résultats de l'exécution sur différentes types de vidéos et interprétation des résultats.

3.2 Implémentation

Pour réaliser notre application nous avons utilisé le langage de programmation orienté objet C++ à cause de sa rapidité dans les calculs (surtout dans le traitement d'images) et la possibilité de rendre le code extensible et réutilisable. L'application a été développée dans l'environnement est le ***Builder Borland v6*** et en utilisant la bibliothèque **VideoLab** sur un micro ordinateur ayant une fréquence de 1,53 GHZ, mémoire vive de 4 Go, disque dur de 500 Go et système d'exploitation Window 8.

La bibliothèque VideoLab : c'est une bibliothèque gratuite qui offre un ensemble de composants pour le traitement vidéo rapide. Elle existe en deux versions une VCL - Delphi / (Daemon Tools sous Windows 7) C + + Builder et une version MFC compatibles Visual C + + de Delphi / C + + Builder.

3.3 Mesure d'évaluation de la méthode de segmentation

Pour l'évaluation de ces techniques de détection de changement de plans, plusieurs critères sont utilisés. Les plus adoptés sont :

Le ''***Rappel***' et la ''***Précision***''. Dans notre cas, l'évaluation consiste à calculer le nombre de plans détectés, le nombre de plan correctement détectés et le nombre de plan en référence. Nous avons utilisé le logiciel AVS Video ReMaker pour détecter le nombre des plans en référence.

Le rappel est le rapport entre le nombre des plans correctement détectées et le nombre des plans en référence.

$$Rappel = \frac{plans\ correctement\ détectés}{plans\ en\ référence}$$

La précision est le rapport entre le nombre des plans correctement détectés et le nombre des plans détectés.

$$\mathbf{Précision} = \frac{plans\ correctement\ détectés}{\mathbf{plans\ détectés}}$$

Les plans détectés sont des plans obtenus par notre algorithme.

- ✓ Les plans correctement détectés sont des plans corrects (par rapport à des plans en référence) détectés par notre algorithme.

3.4. Présentation des vidéos de test

Nous avons travaillé principalement sur six vidéos dans le but d'évaluer et de valider notre système de segmentation en plan de vidéo. Les vidéos utilisées sont de types différents (journal télévisé, film, dessins animés, football, documentaire publicité) avec une variété de contenu (effets spéciaux, sous-titrages, etc.). Ces vidéos sont de type AVI et de fréquence d'image égale à 30 images par seconde. Les séquences vidéo utilisées sont données par la figure 17 :

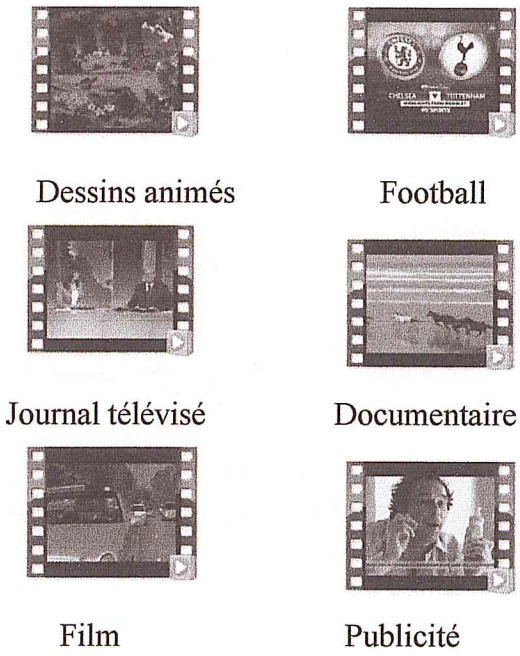


Figure 17 : Listes des Vidéos

Le tableau suivant résume les caractéristiques de six vidéos :

Type de la vidéo	Durée en seconde	Nombre d'images	Nombre de plans en référence	Transitions Cut	Transition Progressive
Film	16	491	7	3	3
Dessin animée	7	208	5	3	1
Football	16	494	4	1	2
Documentaire	9	287	4	3	0
Publicité	11	331	5	4	0
Journal télévisé	8	252	3	1	1

Tableau 1 : Les caractéristiques des vidéos

3.5 Interfaces de l'application

Au lancement de l'application, l'utilisateur dois lire et découper une vidéo sélectionner en séquences d'images, après il choisi la méthode de segmentation ou bien ouvrir le logiciel AVS REMarker pour détecter les plans en référence de la vidéo sélectionner.

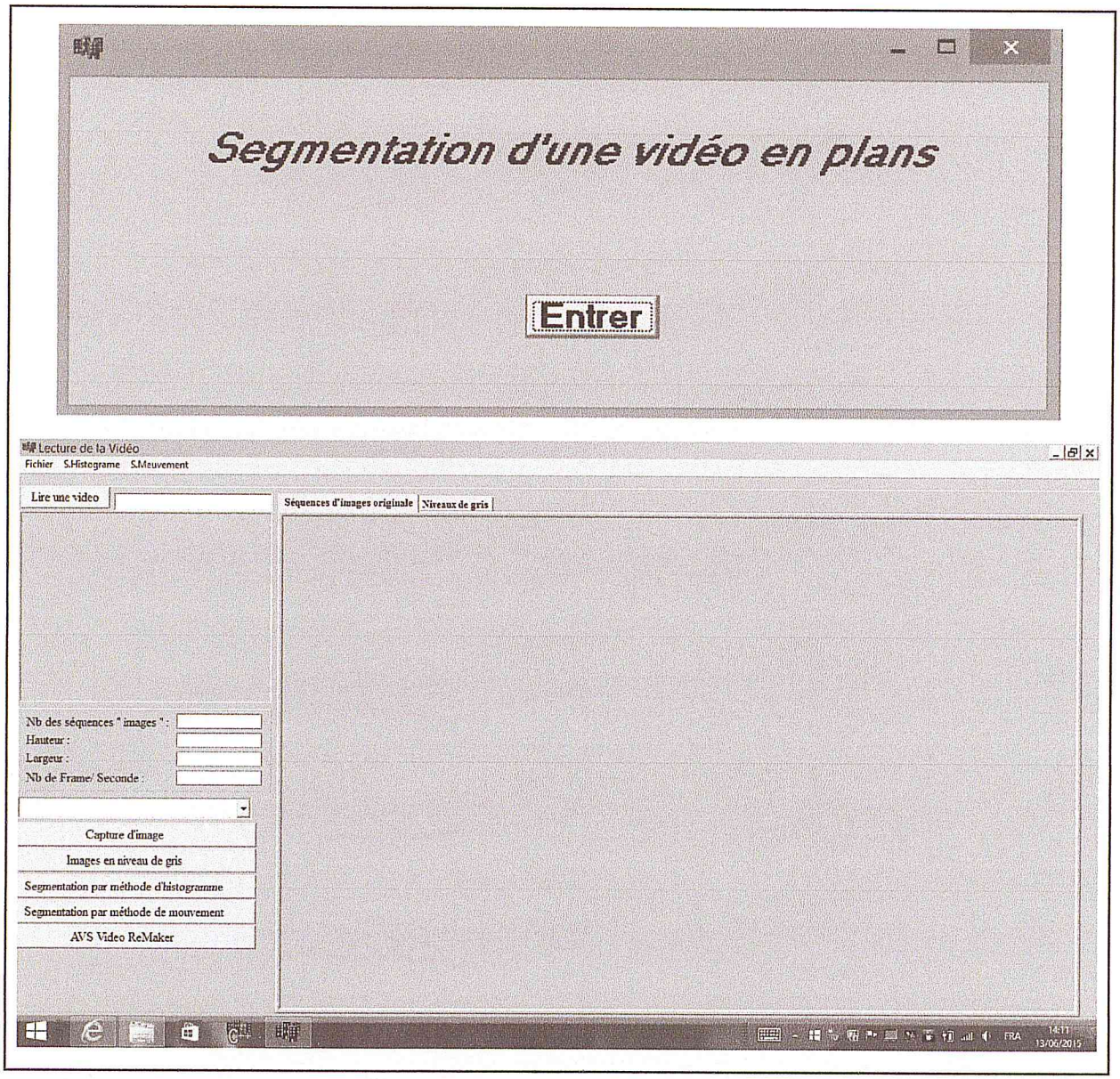


Figure 18 : Entrer de l'application

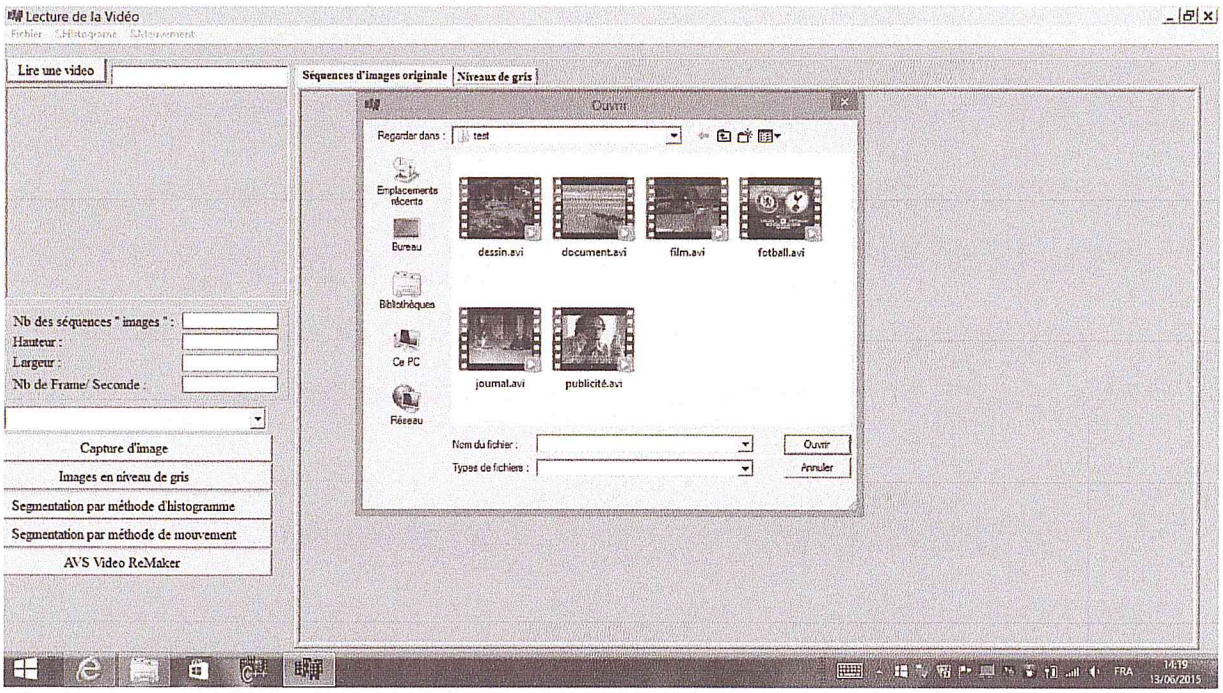


Figure 19 : Sélectionner une vidéo

Figure 20 : Découpage de la vidéo en séquences images

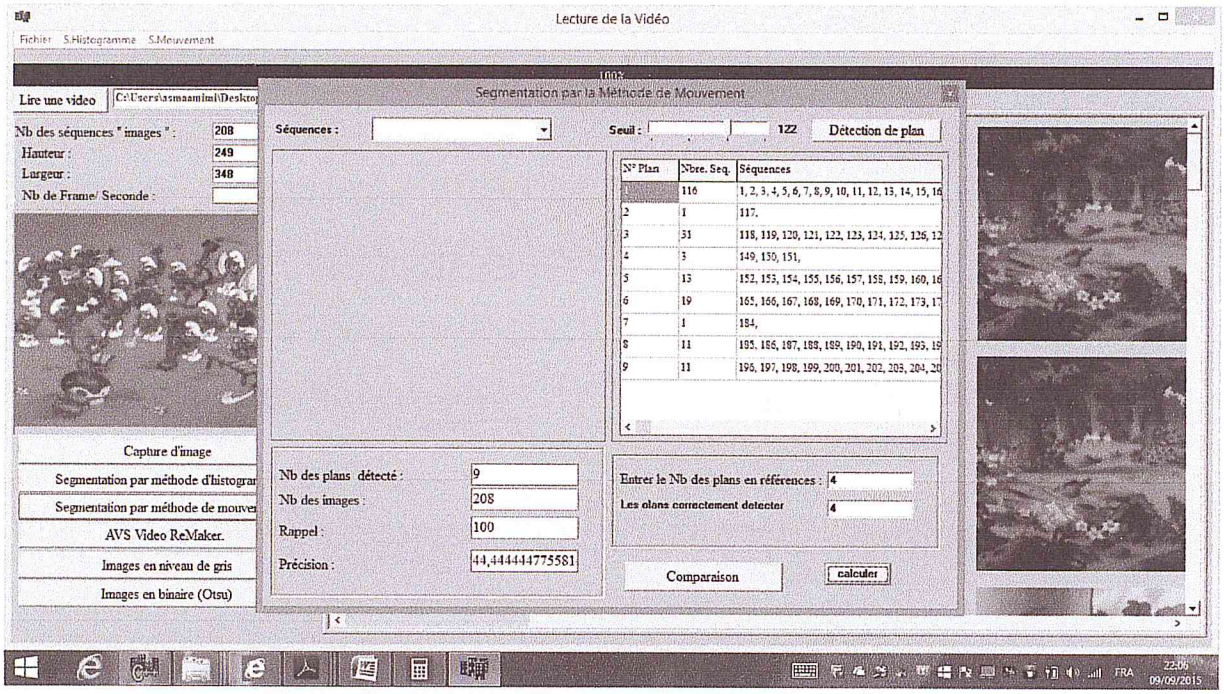


Figure 21 : Résultat d'une méthode de segmentation

3.6. Résultats et discussions

Nous présentons dans cette partie quelques exemples de résultats expérimentaux obtenus par les deux techniques proposées. Ces résultats ont été obtenus en appliquant un seuil d'une façon empirique.

Vidéo	Précision	Rappel
Film	50%	72.72%
Dessin animée	87,5%	76%
Football	33.34%	66 ,66%
Documentaire	100%	100%
Publicité	75%	100%
Journal télévisé	86%	75%

Tableau 2 : Résultats obtenus par histogramme de couleur

A partir de ces résultats, nous pouvons observer que la performance de la méthode histogramme de couleur est bonne dans les vidéos *Dessin animée, Documentaire, Publicité et Journal télévisé* avec un rappel entre 75% et 100% et une précision entre 75% et 100% .Ceci est interprété par la mesure de similarité utilisé (la distance Euclidienne) entre les images successives qui montre bien les types de changement de plans dans ces vidéos(*cut* et *progressive*). Par contre les deux vidéos restent (*Film et Football*) ont moins bon résultats. Ces valeurs diminuent peu les valeurs du rappel et la précision lorsqu'il y a un faible changement du contenu visuel entre deux images (transition progressive).

Vidéo	Rappel	Précision
Film	85,71%	75%
Dessin animée	80%	57,14%
Football	100%	80%
Documentaire	75%	75%
Publicité	60%	60%
Journal télévisé	66,66%	50%

Tableau 3 : Résultats obtenus par mouvement

A partir de ce tableau, nous remarquons que la méthode fournit de bonnes performances avec un rappel entre 85% et 100% et une précision entre 75% et 80% pour les deux vidéos (Film, et Football). Cela est dû à l'activité du mouvement dans la vidéo.

Pour les quatre vidéos restantes, les résultats ne sont pas satisfaisants. Ceci se traduit par la même raison citée auparavant. Nous constatons que la vidéo *Journal télévisé* présente le moins bon résultat. C'est en raison du au contenu des plans de cette vidéo qui correspond à une séquence d'images a moins d'activité. Pour la vidéo *Dessin animés* et *Journal télévisé*, les résultats de la précision sont faibles car l'image disparait progressivement lors d'une transition vers un autre état ; c'est le cas des plans progressifs (fondus enchainés).

3.7 Conclusion

Nous avons présenté dans ce chapitre l'application de segmentation temporelle de vidéos par histogramme de couleur et détection du mouvement. Les résultats obtenus montrent les performances de ces méthodes pour la détection de changement de plans dans le cas des vidéos présentant des changements de plans brusques (*cut*). Par contre, les résultats sont moins performants dans le cas où les vidéos présentent des transitions progressives. Ceci s'explique par les caractéristiques propres pour chaque descripteur utilisé (couleur et mouvement) qui jouent un rôle estimé pour un changement de plan.

Conclusion Générale

L'objectif principal de notre travail consiste à trouver les différents changements des plans de la vidéo par les méthodes de segmentation temporelle. Les techniques utilisées sont basée sur les algorithmes qui permettent d'avoir les plans de la vidéo à partir d'un ensemble d'images obtenu par l'opération de découpage de la vidéo.

Une distance de similarité est définie afin de regrouper des images successives en des unités élémentaires souvent connues sous le nom de ''plans vidéo''.

La procédure de regroupement d'images en plan joue le rôle indispensable et indissociable dans toutes les méthodes de traitement de la vidéo tel que la similitude entre les plans.

La fiabilité des résultats de toute méthode de traitement de la vidéo dépend de la précision de détection des plans. Selon la méthode d'évaluation utilisée '*Rappel et Précision*'.

Les méthodes utilisées montrent en générale des résultats intéressants lorsque le contenu de la vidéo présente des changements de plan de type brusque (*cut*) et moins performantes lorsqu'il y a une transition progressive.

Notre perspective est développée d'autres méthodes de segmentation en plans pour améliorer le domaine de la recherche des vidéos par le contenu visuel, comme la détection des transitions progressives de type *fondu ou volet* par utilisation des descripteurs couleur, texture ,la forme.

Bibliographie

(**) Ce sont les ouvrages que nous avons lu

1. **MERHEB, Maya.** ANALYSE VIDEO ET MACROSEGMENTATION LA MACROSEGMENTATION PAR MATRICE DE SIMILARITE. *DEA d'Informatique.* 2005/2006. (**)

2. **BOUIROUGA, Hajar.** Reconnaissance des scènes vidéo pour adulte. *THÈSE DE DOCTORAT.* 2012. (**)

3. **A., Saoudi.** *Empreinte Numérique et indexation de vidéo.* s.l. : Advestigo, 2007. p. 13. Vol. 1.

4. **A.MAREDJ, F.SAADI,D.MEDDOUR.** *Découpage automatiquede la vidéo en plans.* s.l. : Laboratoire Bases de Données et Systèmes d'information, 2001. (**)

5. **Lefèvre, Sébastien.** Détection d'événements dans une séquence vidéo. [THESE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITE DE TOURS]. 13 décembre 2002. (**)

6. **T.-L.** INDEXATION ET RECHERCHE DE VIDÉO POUR LA VIDÉOSURVEILLANCE. s.l. : Hanoi, 2009. Vol. 1.

7. **BOURENANE, MOHAMED AMINE.** UN OUTIL POUR L'INDEXATION DES VIDÉOS PERSONNELLES PAR LE CONTENU. *MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES.* s.l. : UNIVERSITÉ DU QUÉBEC, AOÛT 2009. (**)

8. **Hu, M.-K.** Visual Pattern Recognition by Moment Invariants. s.l. : IRE Tran sactions on Information Theory, 1962. p. 187.

9. **Reiss, T.** The Revised Fundamental Theorem of Moment Invariants. *Transa ctions on Pattern Anal ysis and Machine Intelligence.* s.l. : IEE, 1991. p. 834.

10. **Athena Vakali, Mohand-Said Hacid, Ahmed Elmagarmid.** MPEG-7 based description schemes for multi-level video content classification Department of Informatics. Aristotle University : 54006 Thessaloniki, May 2003.

11. **Noel E. O'Connor, Edward Cooke, Herve Le Borgne, Michael Blighe, Tomasz Adamek.** the acetoolbox: low-level audiovisual feature extraction for retrieval and classification. Ireland, Dublin City University : Centre for Digital Video Processing.

12. **Joly, M.P Bui Thi and Philippe.** Describing the video: a semiotic approach. Brescia Italia : in Proc. of CBMI , October 2001.

13. **LIU, Dawei.** Indexation d'images par l'histogramme des couleurs. *Rapport de stage du master 2 recherche.* s.l. : Institut des Sciences et Techniques de l'Ingénieur d'Anger, 30 juin 2007. (**)

14. **Imane, DAOUDI.** Recherche par similarité dans les bases de données multimédia : application à la recherche par le contenu d'images. *THÈSE DE DOCTORAT.* AGDAL : UNIVERSITÉ MOHAMMED V, 17 Juillet 2008. (**)

15. **Michael J. Swain, Dana H. Ballard.** Color Indexing. *International Journal of Computer Vision.* 1991. Vol. 7, pp. 11-32.

16. **J.R. Smith, S.F. Chang.** Tools and techniques for color image retrieval. *Storage and retrieval for image and video databases (SPIE).* 1996. pp. 42-64-37.

17. **Kullback, S.** Information theory and statistics. Dover-New York : s.n., 1968.

18. **J. Puzicha, T. Hofman, J. Buhman.** *Non-parametric similarity measures for unsupervised texture segmentation and image retrieval.* s.l. : IEEE Conference on Computer Vision and Pattern Recognition, 1997. pp. 267-272. Vol. .

19. **M. Stricker, M. Orengo.** *Similarity of color images.* 1995. pp. 381-392. SPIE Conference on Storage and Retrieval for Image and Video Databases III.

20. **Mbarek, CHARHAD.** *Modèles de Documents Vidéo basés sur le Formalisme des Graphes Conceptuels pour l'Indexation et la Recherche par le Contenu Sémantique.* 2005. p. 176. (**)

21. **Souvannavong, Fabrice.** Indexation et recherche de plans vidéo par le contenu séemantique. *Thèse présentée pour l'obtention du grade de docteur de Télécom Paris dans la spécialité du traitement du signal et des images.* 03 juin 2005. (**)

22. **Ali, KHALFI.** METHODES D'ARBRES EN INDEXATION ET RECHERCHE D'IMAGES PAR SIMILITUDE VISUELLE. *MEMOIRE DE MAGISTER.* 2015. (**)

23. **Irena Koprinska, Sergio Carrato.** *Temporal video segmentation: A survey.* s.l. : ELSEVIER, 1999. (**)

24. **Kikukawa, T., & Kawafuchi, S.** *Development of an automatic summary editing system for the audio-visual resources.* s.l. : Transactions on Electronics and Information, 1992. pp. 204- 212. Vol. 2.

25. **Zhang, H., Kankanhalli, A, & Smoliar, S. W.** *Automatic partitioning of full-motion video.* s.l. : Multimedia Systems, 1993. pp. 10-28.

26. **Nagasaka, A, & Tanaka, Y.** *Automatic video indexing and full-search for video appearances.* s.l. : Visual database Systems, 1992. pp. 113- 127. Vol. 2.

27. **Alattar, A.M.** *Detecting and compressmg dissolve reglOns m video sequences with a dvi multimedia image compression algorithm.* s.l. : IEEE International Symposium on Circuits and Systems, 1993. pp. 13- 16. Vol. 1.

28. **Gargi, U, Kasturi, R, & Strayer, S.** *Performance characterization of videoshot change detection methods.* s.l. : IEEE Transactions on Circuits and Systems for Video Technology, 2000. pp. 1- 13.

29. **Lienhart, R.** *Comparison of automatic shot boundary detection algorithms.* s.l. : on Storage and Retrieval for Image & Video Databases VII, 1999. pp. 290- 301. Vol. 3656 .

30. **Lupatini, G., Saraceno, c., & Leonardi, R.** *Scene break detection: a comparison. 8th International Workshop on Research Issues in Data Engineering.* 1998. pp. 34-41.

31. **Mediadico.** http ://www.mediadico.coml. [En ligne] [Citation : 15 août 2009.]

32. **Fernando, W.A.C., Canagarajah, C.N., & Bull, D.R.** *Video segmentation and classification for content based storage and retrieval using motion vectors.* 1999. pp. 687-698.

33. **Boreczky, J., & Rowe, L.** *Comparison of video shot boundary detection techniques.* 1996. pp. 170-179. Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases.

34. **Viper** : Multimedia Information Retrieval and Management. [En ligne] 10 août 2009. http://viper.unige.ch/doku.php.

35. **Heesch, D., Howarth, P., Magalhâes, J., May, A., Pickering, M., Yavlinski, A., & S. Rüger**. (2004). *Video Retrieval using Search andiBrowsing. TREC2004 - Text REtrieval Conference,* Gaithersburg, Maryland, 15-19 November

36. **Hadjidemetriou, E., Grossberg, M. D., & and Nayar, S. K**. (2004). Multiresolution Histograms and Their Use for Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence, 26(7):831- 847*

37. **Nayar, S., & Bolle, R.** (1996). Reflectance based object recognition. *Int. J. Comput. Vision, 17(3):219- 240.*

38. **Finlayson, G.D., Chatterjee, S. S., & Funt, B. V.** (1996). Color angular indexing. *Proceedings of the Second European Conference on Computer Vision,* pages 16- 27.

39. **Slater, D., & Healey, G.** (1996). The illumination-invariant recognition of 3d objects using local color invariants. *IEEE Trans. Pattern Anal. Mach. InteU.,* 18(2):206- 210.

40. **Wong, K.-M., Chey, C.-H., Liu, T.-S., & Po, L.M**. (2003). Dominant color image retrieval using merged histogram. *Proceedings of the International Symposium on Circuits and Systems,* Volume 2, pp: II-908 - II-91 1 vol.2, May.

41. **Deng, Y., Manjunath, B.S., Kenney, c., Moore, M.S., & Shin, H.** (2001). An Efficient Color Representation for Image Retrieval. *IEEE Trans. Image Processing,* 10(1): 140- 147.

42. **Kherfi, M.L., Ziou, D., & Bemardi, A.** (2003). Combining positive and negative examples in relevance feedback for content-based image retrieval. *Journal of Visual Communication and Image Representation,* Vol. 14, No. 4. pp. 428-457.

43. **EI-Feghi, I., Aboasha, H., Sid-Ahmed, M.A., & Ahmadi, M.** (2007). ContentBased Image Retrieval based on efficient fuzzy color signature. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics,* pp: 1118 - 1124,October.

44.**Divakaran, A., Ho, H., Sun, H., & Poon, T.** (1998). Scene change detection and feature extraction for indexing MPEG-2 and MPEG-4 sequences. *IEEE Trans. Circuits Systems Video Technology* , Oct.

45. **Fablet, R., Bouthemy, P., & Pérez, P.** (2000). Statistical motion-based video indexing and retrieval. *Proceedings of 6th Int. Conference on Content-Based Multimedia Inf Access,* RIAO'2000, Paris, April, pp. 602-619.

46. **Jain, A.K., Vailaya, A, & Xiong, W.** (1999). Query By Video Clip. *Multimedia Systems: Special Issue on Video Libraries,* vol. 7, no. 5, Mai, pp. 369-384.

47. **Smolic, A, Sikora, T., & Ohm, J.-R.** (1999). Long-term global motion estimation and its application for sprite coding, content description, and segmentation. *IEEE Trans. On Circuits and Systems for Video Technology,* Vol. 9, No. 8, December, pp. 1227-12242.

48. **Gelgon, M., & Bouthemy, P.** (1998). Determining a structured spatio-temporal representation of video content for efficient visualization and indexing. *Proceedings of the 5th European Conference on Computer Vision,* ECCV'98, Springer, Freibourg, Juin, Vol 1406, pp. 595-609.

49. **Jeannin, S., Mory, B.** (2000). Video Motion Representation for Improved Content Access. *IEEE Trans. on Consumer Electronics,* Vol. 46, No. 3, August, pp. 645-655.

50. **Deng, Y., & Manjunath, B.S.** (1998). NeTra-V: Toward an Object-based Video Representation. *IEEE Transaction on Circuits and Systems for Video Technology,* Vol. 8, No.5, Septembre, pp. 616-627.

51. **Zaharia, T., Prêteux, F.** (1999). *Motion descriptor: perspective transformation parameters and object trajectory.* Proposal P351, MPEG-7 Proposai Evaluation Meeting, Lancaster, UK, Feb.

52. **Prêteux, F., Zaharia, T., & Pre da, M.** (1999). Parametric Object Motion Descriptor. *ISO/IEC JTCi/SC29/WGll, MPEG99/M4870,* Vancouver, BC, Canada, July.

53. **Zaharia, T., & Prêteux, F**. (2001). Parametric motion models for video content description within the MPEG-7 framework. *Proceedings SPIE Conf on Nonlinear image Proc. and Pattern Analysis,* San Jose, USA, 22-23 January.

54. **Chang, S.-F., Chen, W., Meng, H.J., Sundaram, H., & Zhong, D**. (1998). A Fully Automated Content-Based Video Search Engine Supporting Spatiotemporal Queries. *iEEE Transactions on Circuits and Systems for Video Technology,* Vol. 8, No. 5, September, pp.602-615.

55. **HADI, M. Rachid OULAD HAJ THAMI M. Youssef.** Développement d'une application de segmentation de la vidéo en se basant sur l'hisogramme de couleur. *Fin d'étude 2ème année.* 2005 - 2006.

56. **H. Zhang, A. Kankanhalli, and S. Smoliar**, *Automatic Partitioning of Video*, Multimedia Systems, Volume 1, Number 1, 1993, pp. 10-28.

57. **Youssef HADI, Fedwa ESSANNOUNI, Rachid OULAD HAJ THAMI, Ahmed SALAM and Driss ABOUTAJDINE**, *A New Approach for Video Cut Detection Using Color Histogram*, ISIVC 2006, Hammat Tunis, 2006.

58. 1. **ChristopherLizant.** Binarisation d'image Methode d'OTSU. [En ligne] [Citation : 12 08 2014.] https://sites.google.com/site/lizantchristopher/services/binarisation-1. (**)

59. **H. BOUIROUGA, A. JILBAB, D. ABOUTAJDINE**,"Reconnaissance des scènes vidéo adultes", Mémoire de fin d'études, Novembre 2007. (**)