MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE YAHIA FARES DE MEDEA
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES ET D'INFORMATIQUE
FILIERE : INFORMATIQUE



MEMOIRE DE FIN D'ETUDES

EN VUE DE L'OBTENTION DU DIPLOME DE MASTER EN INFORMATIQUE

Option :Ingénierie des systèmes et technologie du web

Détection des points de transition pour la segmentation et structuration des vidéos de conférance

Présenté par :
Aya FEKIR
Karima BENLEKEHAL

Dirigé par :
Dr.BELKACEM Soundes

Année Universitaire 2020/2021

Remerciement

On remercie le Dieu le tout puissant de nous avoir donnée la volonté et la sante d'entamée et termine ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de DR. Soundes BELKACEM, on la remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

Nous tenons également à remercier les membres de jury d'avoir accepté d'examiner ce travail.

Nos remercient s'adresse également à tous nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Dédicace

Je dédié ce mémoire a :

Ma chère mère, Autant de phrases aussi expressives soient-elles ne sauraient montrer le degré d'amour et d'affection que j'éprouve pour toi. Tu n'as cessé de me soutenir et de m'encourage durant toutes les années de mes études, En ce jour mémorable, pour moi ainsi que pour toi, reçoit ce travail en signe de ma vive reconnaissance et ma profonde estime. Puisse le tout puissant te donner santé, bonheur et longue vie afin que je puisse te combler à mon tour.

Mon cher père, Tes conseils ont toujours guidé mes pas vers la réussite. Ta patience sans fin, ta compréhension et ton encouragement sont pour moi le soutien indispensable que tu as toujours su m'apporter. Je te dois ce que je suis aujourd'hui et ce que je serai demain et je ferai toujours de mon mieux pour rester ta fierté et ne jamais te décevoir. Que Dieu le tout puissant te préserve, t'accorde santé, bonheur, quiétude de l'esprit et te protège de tout mal.

Ma chère sœur et mes chers frères vous êtes toujours ma force, ma source d'espoir que Dieu vous protège

Ma chère tante Farida qui n'est pas là pour moi aujourd'hui, qui reste toujours vivante dans mon cœur, qu'ALLAH nous rassemble a Jannah AL-FIRDAWS

Mes chères grands-mères Houria et Meriem vos douaa sont toujours le compagnon en détresse, puisse le dieu le tout puissant vous donner sante, chiffa et longue de vie.

Me chère famille qui ma arrosés de tendresse d'espoir, d'amour Merci infiniment

Ma partenaire dans le projet Karima, tu as le meilleur cœur du monde, j'apprécie tous ce que tu as fait pour moi ce que soit amour, motivation, courage. Tu es la douce force.

Mes meilleurs amies Maria et Chaimaa Ma promotrice Dr. Soundes BELKACEM qui a tout fait pour nous encouragées, nous motivées.

aya FEKIR

Dédicace

Je dédié ce modeste travail accompagne d'un profond amour :

A la source d'amour incessible, a la mère des sentiments fragiles qui ma bénie par ces prières et douaa.

A mon support dans la vie, a qu'il m'a appris à supporter et ma diriger vers la galère, mon père, Que Dieu le tout puissant te préserve et te protège pour moi.

Ma chère sœur Sarah et mon chère frère Youcef qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité, que dieu vous offre la chance le bonheur

Ma chère Fatima qui a étais avec moi des mes premiers pas, qui as tout fait pour moi Mon cher oncle, mon deuxième papa tu restes toujours dans mon cœur, qu'ALLAH nous Rassemble a Jannah AL-FIRDAWS Me chère famille qui m'a encouragé motivée.

Ma promotrice Dr. Soundes BELKACEM , j'apprécie tous ce que tu as fait pour nous ce que soit motivation, courage, et espoir

Benlakehel Karima

Abstract

Nowadays, conference videos have become an essential medium for knowledge sharing and learning due to their great ability to provide rich information. This information must be indexed by a conference video search system. Video temporal segmentation system is an essential step for videos indexing. It makes easier to navigate within conference videos contents and retrieve important information. In general, temporal segmentation system consists of three main phases: features extraction, similarity measurement and detection. The aim of this system is to detect the transition points between shots in a conference video. However, they show significant changes in illumination, rapid objects or camera movement which makes detection task complex and difficult to perform. In this work, feature extraction phases require frames prepossessing which consists of : haze removal and color space conversion into the CIELAB space. Enhanced frames descriptors will be analyzed and will be used for similarity measurement based on the CIEDE2000 formula. Finally, the detection phase is based on a test with global threshold. For evaluation, we use a dataset of conference videos consisting of two videos types: type 1 for videos containing only one slide view and type 2 of videos containing both slide and speaker view. The experimental results prove proposed system efficiency compared to some existing methods.

Keywords : conference videos, video indexing, video temporal segmentation, transition points, slide transitions, CIEDE2000, global threshold.

Résumé

De nos jours, les vidéos de conférences sont devenues un support essentiel pour le partage des connaissances et l'apprentissage en raison de leur grande capacité à fournir des informations riches. Ces informations doivent être indexées par un système de recherche des vidéos de conférence. Le système de segmentation temporelle des vidéos est une étape d'indexation des vidéos. Cette dernière, permet de faciliter la navigation des vidéos de conférence et la récupération des informations importantes. En général, le système de segmentation temporelle des vidéos de conférence consiste de trois phases principales : l'extraction des caractéristiques des plans, mesure de similarité et la détection. L'objectif de ce système a détecté les points de transition entre les plans d'une vidéo de conférence. Mais, les vidéos de conférence présentent plusieurs changements de lumination, des mouvements rapides d'objets ou de caméras ce qui rend la tâche de détection très complexe et difficile à réaliser. Dans ce travail, la phase d'extraction des caractéristiques nécessite le prétraitement des plans, cette dernière se base sur l'amélioration qui consiste : la suppression de la brume des plans et la transformation d'espace colorimétrique vers l'espace CIELAB, les plans améliorés vont être analysés pour l'extraction des descripteurs qui seront utilisés dans la phase de masure de similarité, cette dernière, se base sur la formule CIEDE2000. Enfin, la phase de détection se base sur un test avec seuil global. Pour l'évaluation on utilise une base des vidéos de conférence de deux types : type 1 des vidéos contient une vue de diapositive seulement et type 2 des vidéos contient une diapositive et vue d'interlocuteur. Les résultats d'expérimentation prouvent l'efficacité du système par rapport à des méthodes existantes.

Mots clé : les vidéos de conférence, indexation des vidéos, segmentation temporelle des vidéos, les points de transition, les transitions de diapositive, CIEDE2000,seuil global.

الملخص

في الوقت الحاضر ، أصبحت مقاطع الفيديو الخاصة بالمؤتمرات وسيلة أساسية لمشاركة المعرفة والتعلم نظرًا لقدرتها الكبيرة على توفير معلومات ثرية. يجب فهرسة هذه المعلومات بواسطة نظام بحث فيديو المؤتمر. يعد نظام التقسيم الزمني للفيديو خطوة لفهرسة مقاطع الفيديو. هذا الأخير يجعل من السهل التنقل في مقاطع الفيديو الخاصة بالمؤتمر واسترداد المعلومات المهمة. بشكل عام ، يتكون نظام التقسيم الزمني لمقاطع فيديو المؤتمر من ثلاث مراحل رئيسية: استخراج الميزات من اللقطات وقياس التشابه والكشف. كشفت عدسة هذا النظام عن نقاط الانتقال بين المقاطع في مؤتمر الفيديو. ولكن ، تُظهر مقاطع الفيديو الخاصة بالمؤتمر تغييرات متعددة في الإضاءة ، والحركة السريعة للأشياء أو الكاميرات ، مما يجعل مهمة الكشف معقدة للغاية ويصعب تنفيذها. في هذا العمل ، تتطلب مرحلة استخراج الميزات المعالجة المسبقة للخطط ، ويستند الأخير إلى التحسين الذي يتكون: إزالة ضباب الخطط وتحويل مساحة CIELAB اللون إلى مساحة أخيرًا ، تعتمد مرحلة الاكتشاف على اختبار بحد إجمالي. للتقييم ، CIEDE2000. مرحلة التشابه ، وتستند الأخيرة إلى صيغة نستخدم قاعدة من مقاطع الفيديو الخاصة بالمؤتمر من نوعين: النوع 1 من مقاطع الفيديو يحتوي على عرض شرائح فقط والنوع 2 من مقاطع الفيديو يحتوي على عرض شريحة وعرض مكبر صوت. أثبتت النتائج التجريبية فعالية النظام مقارنة بالطرق الحالية.

الكلمات الرئيسية: مقاطع فيديو المؤتمر ، فهرسة الفيديو ، تقسيم وقت الفيديو ، نقاط الانتقال ، انتقالات الشرائح ،، العتبة العالمية Ciede2000.

TABLE DES MATIÈRES

Introduction Générale **1**
Contexte et problématique ... 1
But et objectifs ... 2
Structure du mémoire ... 2

1 Les vidéos pour l'apprentissage en ligne **3**
1.1 Introduction ... 4
1.2 L'apprentissage en ligne ... 4
1.2.1 Apprentissage asynchrone ... 4
1.2.2 Apprentissage Synchrone ... 5
1.2.3 L'utilité des vidéos de conférence ... 5
1.3 Les vidéos ... 6
1.3.1 Définition d'une vidéo ... 6
1.3.2 Structure d'une vidéo ... 6
1.3.2.1 Plan ... 7
1.3.2.2 Prise de vue ... 7
1.3.2.3 Les scènes ... 7
1.3.2.4 Les objets ... 7
1.4 Les caractéristiques d'un doucement vidéo ... 8
1.4.1 Caractéristiques physiques ... 8
1.4.2 Caractéristiques visuelles ... 9
1.4.2.1 La couleur ... 10
1.4.2.2 La texture ... 11
1.4.2.3 La forme ... 11
1.4.2.4 Le mouvement ... 12
1.5 L'indexation des vidéos ... 12
1.5.1 Indexation des vidéos de conférence ... 12
1.5.2 La segmentation temporelle ... 13
1.6 Conclusion ... 14

2 Les systèmes de segmentation temporelle d'un document vidéo **15**
2.1 Introduction ... 16
2.2 Les transitions dans les vidéos ... 16
2.2.1 Définition d'une transition ... 16
2.2.2 Les types de transition d'une vidéo ... 16

2.2.2.1 Transitions dures 17
2.2.2.2 Transitions douces 17
2.2.3 Transition de diapositive 17
2.3 Architecture générale d'un système de segmentation temporelle 18
2.3.1 Extraction des caractéristiques 19
2.3.2 Mesures de similarité 19
2.3.3 Détection 19
2.4 Les méthodes de segmentation temporelle des vidéos 19
2.4.1 Amélioration 19
2.4.1.1 La suppression de la brume 20
2.4.1.2 Le double éclairage 20
2.4.2 Représentation des informations visuelles 20
2.4.2.1 Méthode basée sur les pixels 20
2.4.2.2 Méthode basée sur l'histogramme 21
2.4.2.3 Méthode basée sur les conteurs 21
2.4.3 Construction du signal de dissemblance/similarité 22
2.4.3.1 La formule de différence de couleur CIEDE 2000 22
2.4.3.2 Les avantages du CIEDE2000 22
2.4.4 Détection 23
2.4.4.1 Estimation du seuil 23
2.4.4.2 Classifications 23
2.5 Métriques d'évaluation d'un système de segmentation temporelle 23
2.6 Conclusion 24

3 Conception **25**
3.1 Introduction 26
3.2 Architecture générale du système 26
3.3 Description détaillée 27
3.3.1 Phase 1 : le prétraitement 27
3.3.1.1 Suppression de la brume des plans 27
3.3.1.2 Transformation d'espace de couleurs 29
3.3.2 Phase 2 : L'analyse 29
3.3.2.1 Extraction des caractéristiques 29
3.3.2.2 Calcul de la distance CIEDE2000 30
3.3.3 Phase 3 : La détection des points de transition 31
3.4 Conclusion 32

4 Résultats et discussions **34**
4.1 Introduction 35
4.2 Environnement de développement 35
4.2.1 Environnement matériel 35
4.2.2 Environnement logiciel 35
4.3 Base de données utilisée 35
4.3.1 La collection des vidéos 35
4.3.1.1 Vidéos de type 1 36
4.3.1.2 Vidéos de type 2 36
4.3.2 L'annotation de la base 37
4.4 Métriques d'évaluation 37
4.5 Résultats et discussions 38

4.5.1 Paramètres de configuration....38
4.5.1.1 La phase de prétraitement....38
4.5.1.2 La phase d'analyse....38
4.5.1.3 La phase de détection....38
4.5.2 Evaluation et discussions....40
4.5.2.1 Phase de détection....40
4.5.3 Comparaison avec d'autres méthodes....41
4.5.3.1 Comparaison avec autres méthodes d'amélioration....41
4.5.3.2 Comparaison avec autres méthodes de détection....41
4.5.3.3 Comparaison en différents espaces de couleur....42
4.6 Conclusion....42

5 Conclusions générale **43**
5.1 Conclusion....44
5.2 Perspectives....44

TABLE DES FIGURES

1.1 Échantillon d'un signal vidéo [17] 6
1.2 Structure hiérarchique d'une vidéo 7
1.3 Modèle de couleur HSV [26] 10
1.4 Modèle de couleur RGB [27] 11
1.5 Modèle l'espace colorimétrique CIELAB [28] 11
1.6 Exemple des deux types de vidéos de conférence 12
1.7 Illustration du processus de la segmentation temporelle 13

2.1 Période d'une transition entre les plans [1] 16
2.2 Les types de transitions d'une vidéo [1] 16
2.3 Exemple d'une transition dure entre les deux plans 1952,1953.[1] 17
2.4 Exemple de transition fondu entrant avec 21 plans impliqués dans la transition (n=22-1) 18
2.5 Exemple de transitions diapositive (A) vidéo de type 1 (B) vidéo de type 2. 18
2.6 Architecture générale de système temporelle des vidéos 19
2.7 Les étapes de calcul du CIEDE2000 22

3.1 Architecture générale du système 26
3.2 Description générale de la méthode de suppression de brume 27
3.3 Exemples de résultat de suppression du brume(A) plan original, (B)plan après la suppression du brume 28
3.4 Transformation d'espace de couleurs, (a) image dans l'espace BVR(b) image dans l'espace CILAB 29
3.5 Exemple de la division des plans : (A) plan original, (a) sous plan 1,(b) sous plan 2, (c)sous plan 3 et (d) sous plan 4. 30
3.6 le processus de CIEDE2000 entre plan (A) et le plan (B) 31
3.7 Organigramme de la phase de détection 32
3.8 Un exemple des résultats de la phase de détection : détection d'une transition de diapositive entre les deux plans consécutifs 3723,37234 33

4.1 Exemples des plans des vidéos de la base de données (a) vidéo de type 1 (b) et (c)vidéo de type 2 36
4.2 Mesure F1 pour différents seuils 38
4.3 Exemples de résultats de suppression du brume(A) plan original, (B)plan après la suppression du brume 39

4.4 Comparaison des performances de détection sur la vidéo 1 pour différents nombres de sous-plans 40
4.5 Comparaison des performances de détection sur la vidéo 9 pour différents nombres de sous-plans 40
4.6 La différence CIEDE2000 entre plans de la vidéo 1 40
4.7 La différence CIEDE2000 entre plans de la vidéo 11 40
4.8 Evaluation des performances sur les vidéos de type 1 41
4.9 Evaluation des performances sur les vidéos de type 2 41

LISTE DES TABLEAUX

1.1 exemple d'un catalogue contenant des clés physiques d'un fichier vidéo. 9

4.1 Caractéristiques des vidéos du type 1. 36
4.2 Caractéristiques des vidéos du type 2. 37
4.3 Les moyennes de précision, de rappel et de F1-mesure pour chaque type de vidéos. 41
4.4 Example sur les valeurs de performances de système entre la méthode de suppression de la brume et la méthode de double éclairage pour le vidéo 1 41
4.5 Les moyennes de précision, de rappel et de F1-mesure des vidéos de Type-1 pour chaque méthode 42
4.6 Les moyennes de précision, de rappel et de F1-mesure des vidéos de Type-2 pour chaque méthode 42
4.7 Etat comparatif de la méthode proposée en espaces de couleur RVB et LAB pour le type 1 42
4.8 Etat comparatif de la méthode proposée en espaces de couleur RVB et LAB pour le type 2 42

INTRODUCTION GÉNÉRALE

Avec la distanciation sociale que nous impose le virus Corona, nous avons besoin de bonnes technologies pour fournir les meilleures plates-formes de communication à distance grâce à Internet, nous pouvons accéder à plusieurs informations à tout moment et en tout lieu, ce qui a conduit à l'expansion de l'utilisation d'Internet et l'augmentation considérable de la demande d'information et de leurs disponibilités. La création de données sur Internet est illimitée, ce qui rend difficile l'accès efficace à ces informations. Parmi les données les plus utilisées et plus mises en ligne sont : vidéos à usage éducatives, spécialement les vidéos de conférences appelées aussi les vidéos de présentations, qui sont devenues un support essentiel pour le partage des connaissances et l'apprentissage en raison de leur grande capacité à fournir des informations riches d'un point et de leur production facile à base de technologies communément disponibles tel que les téléphones portables. Les utilisateurs peuvent accéder librement à de nombreuses vidéos via Internet pour acquérir des connaissances. Cependant, ils ont souvent difficulté à trouver des informations spécifiques à partir des vidéos. Par exemple, lorsque les étudiants souhaitent revoir une certaine partie d'un cours, ils peuvent être obligés de parcourir la quasi-totalité de la vidéo, ou la lire plusieurs fois pour trouver la bonne position [12]. Pour résoudre ce problème, de nombreuses recherches ont été menées pour indexer automatiquement le contenu des vidéos de conférence. Pour permettre l'indexation, une vidéo doit être divisée en segments disjoints significatifs, appelés pris de vue ou segments. Pour les vidéos de conférences, la segmentation des prises de vues est généralement déterminée en fonction des points où se produit une transition de diapositive, car la plupart des cours sont présentés sous forme de diaporamas. Dans ce cas, une prise de vues peut être considérée comme une séquence de plans (éléments d'image d'une vidéo) appartenant à la même diapositive. Le point de transition de la diapositive devient une limite de prise de vues [30].

Contexte et problématique

L'indexation automatique des vidéos de conférence nécessitent que leurs contenues doit être en premier organisé et analysé complètement par des systèmes de segmentation temporelle de vidéos. Les systèmes de segmentation temporelle de vidéos permettent de détecter les déférentes transitions entre les prises de vues des vidéos. Généralement, ces systèmes sont composés de trois étapes : extraction des caractéristiques visuelles, mesure de similarité et la détection.

La phase d'extraction des caractéristiques a pour objectif l'extraction efficace des caractéristiques visuelles à partir des plans vidéo. Les performances de l'extraction des caractéristiques sont aussi liées à la résolution et la qualité des vidéos spécialement les vidéos

de conférence , ces derniers ne seront pas toujours de haute qualité en raison de nombreux facteurs défavorables, tels que la variété des types de vidéo, les effets spéciaux et le bruit produit par les changements d'illumination ou mouvements d'objets ou de caméras et l'utilisation d'équipements simples pour le montage des vidéos. Ces dégradations peuvent mener à des détections d'erreurs dans la phase de détection. La construction du signal de dissemblance/similarité est affectée par la qualité de description.

Un système de segmentation temporelle des vidéos robuste devrait effectuer de bonnes détections des transitions pour tous les types de transition mais aussi moins de fausses détections ou d'omissions. Dans ce travail, nous soulevons les problématiques suivantes : comment effectuer une extraction des caractéristiques en remplissant deux exigences : invariance et robustesses par rapport aux différentes dégradations ? Comment mesurer la dissimilaire entre les plans pour la détection des transitions de diapositive dans les vidéos de conférence et une fois détecter ? Età quelle seconde ils apparaissent dans la vidéo.

But et objectifs

L'objectif de ce travail est de proposer et implémenter un système de détection des points de transition de diapositive dans les vidéos de conférences qui permet la navigation et l'accès à leurs contenus. Afin d'atteindre notre objectif principal, d'autres sous objectifs spécifiques ont étais spécifiés :

1. Présenter une étude approfondie sur les documents vidéo ;
2. Présenter un état de l'art et comparatif sur les méthodes de détection de points de transition ;
3. Conception et implémentation d'une approche de détection des points de transition ;
4. Évaluation de l'approche par des expérimentations sur des vidéos de conférences.

Organisation du mémoire

Ce mémoire, s'organise en quatre chapitres : dans le premier chapitre, nous présenterons l'importance des vidéos de conférence dans l'apprentissage en ligne ainsi que leur segmentation est indexation. Ensuite, le deuxième chapitre met le point sur l'architecture des systèmes de segmentation temporelle et nous verrons les différentes étapes et méthodes utilisées dans ces systèmes. Dans le troisième chapitre, nous décrierons en détails notre système de détection des transitions ainsi les techniques et méthodes utilisées. Le quatrième chapitre présente l'implémentation des méthodes, les détails matériels, logiciels, paramètres de configuration et description des bases de données que nous avons utilisée. Aussi, nous présenterons en détail les résultats des expérimentations de notre systeme avec discussions. En dernier, on conclut notre travail par une conclusion générale et en envisager des futures perspectives.

CHAPITRE 1

LES VIDEOS POUR L'APPRENTISSAGE EN LIGNE

Sommaire

1.1 Introduction 4
1.2 L'apprentissage en ligne 4
1.2.1 Apprentissage asynchrone 4
1.2.2 Apprentissage Synchrone 5
1.2.3 L'utilité des vidéos de conférence 5
1.3 Les vidéos 6
1.3.1 Définition d'une vidéo 6
1.3.2 Structure d'une vidéo 6
1.4 Les caractéristiques d'un doucement vidéo 8
1.4.1 Caractéristiques physiques 8
1.4.2 Caractéristiques visuelles 9
1.5 L'indexation des vidéos 12
1.5.1 Indexation des vidéos de conférence 12
1.5.2 La segmentation temporelle 13
1.6 Conclusion 14

1.1 Introduction

À l'apparition de l'apprentissage en ligne, les ressources pédagogiques distantes consistaient souvent des informations textuelles raison pour laquelle l'apprentissage en ligne n'était pas assez efficace ni populaire chez les apprenants. Les avancées technologiques permettaient d'ajouter du continue multimédia et interactif, ainsi, créer une vision différente de l'enseignement, en offrant des outils permettant des apprentissages différents selon les étudiants [3] et l'enseignant assurait que l'information est accédée de manière correcte et qu'elle parvient aux étudiants avecune haute et bonne qualité. Parmi ces techniques disponibles sont : audio, vidéo et texte. Dans ce chapitre nous présenterons les techniques d'apprentissage ainsi que les caractéristiques des documents vidéo.

1.2 L'apprentissage en ligne

Le e-learning ou encore appelé « formation en ligne », « apprentissage en ligne ». [2,6] est une technique d'apprentissage en ligne

1. **Définition1** : la définition donnée par la commission européenne : utilisation des nouveaux techniques multimédias et de l'internet pour améliorer la qualité de l'apprentissage en facilitant l'accès à des ressources et des services, ainsi que les échanges et la collaboration à distance. Cette définition qui semble être claire, synthétique et complète, met l'accent sur la qualité d'apprentissage et tente d'influer les usages et améliorez les pratiques [1,3].
2. **Définition2** : La définition proposée par le Lab Set : E-Learning (ou Electronic Learning) : apprentissage en ligne centré sur le développement de compétences par l'apprenante et structurée par les interactions avec le tuteur et les pairs [4].

L'apprentissage en ligne réunit deux domaines : l'apprentissage et la technique : il s'agit d'acquérir de connaissances en utilisant technologie comme un outil dans la pratique éducative. Les systèmes d'apprentissage en ligne combinent divers outils, tels que les techniques d'écriture, les techniques de communication, la visualisation et le stockage [5].

1.2.1 Apprentissage asynchrone

Permets aux apprenants à leur propre rythme, sans interaction directe avec l'enseignant via le web. Il s'agit d'informations accessibles en auto-assistance aux apprenants et fournit aussi des moyens d'interaction entre les apprenants à travers la messagerie ou les espaces de discussions [7,8]. Parmi les outils de l'apprentissage asynchrone on trouve :

1. **Le texte** Les informations textuelles sont simples à modéliser au sein des systèmes informatiques [9]. Dans l'apprentissage en ligne les textes représentés par des logiciels Powerpoint, PDF ou WORD peuvent suffire pour créer des ressources d'apprentissage simples, comme une présentation ou un tutoriel. Cependant, des outils plus sophistiqués sont nécessaires pour créer des contenus interactifs. Les logiciels de création du contenu pédagogique sont des outils spéciaux qui permettent de créer des contenus d'apprentissage interactifs. Ils permettent d'insérer du texte, des illustrations et d'autres composantes multimédias, tout en fournissant également un cadre pour organiser les pages et les leçons afin de faciliter la navigation. Les organisations et les établissements d'enseignement recourent de plus en plus à des plateformes d'apprentissage pour dispenser des cours aux apprenants et gérer leurs activités en ligne. Une plateforme d'apprentissage est un ensemble de services interactifs en ligne qui offre aux apprenants un accès à des informations, des outils et des ressources qui favorisent la gestion et

La diffusion de la formation. Ils fournissent un accès et des services à une large base d'utilisateurs via l'internet [10].

2. **L'audio** Le son est une donnée temporelle qui peut évoluer vite, il porte une grande quantité d'informations [11] les conférences audios dans l'apprentissage aient lieu en utilisant les écouteurs et le microphone par les étudie et les ce qui leur permet à la fois de s'écouter et de se parler, mais ne pas se voir. En plus des installations audios, un espace peut être ouvert pour des entrées écrites supplémentaires pour compléter la conférence audio et fournis un nombre défini de sous conférences (appelées « salles ») au sein de chaque conférence, les conférences audios permettent aux apprenants de créer leurs propres salles virtuelles où ils peuvent se rencontrer en paires ou en petits groupes pour participer en temps réel. Les conférences audios sont recommandées pour certains domaines comme l'apprentissage des langues [12].
3. **Les vidéos** Sont plus populaires parmi les ressources d'apprentissage en ligne, car elles peuvent être visualisées, mises en pause et rejouées si nécessaire à la convenance des utilisateurs. De nos jours, les universités et les établissements d'enseignement produisent une vaste collection de vidéos de conférences et les publient en ligne pour que les utilisateurs puissent les visualiser indépendamment de l'heure et de leur emplacement [13]. Les vidéos de conférences ont deux types de sources d'information, l'audio ou à partir du texte en vidéo. Pour extraire les informations disponibles dans les vidéos de conférence, différentes méthodes sont disponibles, pour la récupération d'informations audio par reconnaissance vocale automatique (ASR) et pour les informations textuelles, la reconnaissance optique de caractères (OCR) peut être utilisée. La technique ASR extrait la parole ou la voix de l'audio et la convertit en informations textuelles. Les informations récupérées à l'aide de l'ASR dépendent de la fluidité de la parole du locuteur. Il existe différentes méthodes d'indexation et de recherche de vidéo de conférence qui utilisent le texte extrait de la vidéo de conférence, l'audio extrait et à la fois du texte et de l'audio [14].

1.2.2 Apprentissage Synchrone

Les participants interagissent avec l'enseignant sur le Web en temps réel. Par exemple : des équipements vidéo. Les étudient interagissent avec les enseignants via la messagerie instantanée, le chat, les conférences audio et vidéo, etc. De plus, toutes les sessions peuvent être enregistrées et lues. [7,8]

1.2.3 L'utilité des vidéos de conférence

La vidéo de conférence est souvent utilisée pour l'apprentissage à distance, mettant en relation des enseignants et des apprenants distants pour la prestation de l'instruction. Certaines des façons dont la technique est utilisée pour atteindre cet objectif, apportant des avantages aux enseignants et les étudiants sont les suivants : [15]

- Les étudiants peuvent suivre des cours non offerts dans leur établissement ;
- Les établissements d'enseignement peuvent offrir des cours pendant les heures creuses et aux étudiants qui ne peuvent pas assister aux cours traditionnels.
- Les enseignants peuvent enseigner en équipe avec des enseignants à distance, en partageant leur expertise en la matière ou une approche unique d'un sujet.
- Les « maîtres enseignants » dans des matières particulières peuvent être mis à la disposition d'un public plus large d'étudiants de portée géographique plus large.
- Communication directement avec des experts dans de nombreux domaines pour améliorer la compréhension d'un sujet qu'ils étudient.

— Les enseignants et les étudiants peuvent collaborer et échanger des informations avec d'autres écoles dans des domaines tels que le conseil par les pairs.
— La vidéo de conférence peut faciliter l'apprentissage coopératif distribué, où des groupes situés sur des sites distants entreprennent une tâche d'apprentissage et enseigner à des pairs distants.
— Les projets distribués peuvent utiliser la technologie de vidéoconférence pour la collabo - ration et la communication.
— Les vidéos de conférence facilitent la mémorisation et l'assimilation.

1.3 Les vidéos

1.3.1 Définition d'une vidéo

Les vidéos contiennent des informations plus riches que les images, Une vidéo est un signal composé d'une séquence d'images avec une fréquence d'images spécifique (Frate : mesurée en image par seconde ou fps) accompagnée d'une piste audio. Les vidéos définies comme un signal 3D dans lequel [16] :

— L'axe horizontal (Nx) est la largeur de plan.
— L'axe vertical (Ny) est la hauteur de plan représentant le contenu visuel.
— Le troisième axe représente la variation du contenu de plan avec le temps (mesuré en T $\times$ Nf tel que T est le temps total de la vidéo en secondes ou Nf le nombre de plans de vidéo).

Par conséquent, un point dans une vidéo est identifié par ses positions 2D : (positions de pixels x et y) et le temps ou l'index d'image auquel il se produit. Comme le montre la figure suivante [17] :



FIGURE 1.1 – Échantillon d'un signal vidéo [17]

1.3.2 Structure d'une vidéo

Une vidéo peut être vue comme une séquence d'images bidimensionnelles, L'image 2D est une représentation graphique à base de points colorés appelés des pixels dans un espace à deux dimensions (x, y). La résolution de la vidéo, exprimée en nombre de pixels.

La structure d'une vidéo est représentée par les paramètres suivants :

1.3.2.1 Plan

Un plan d'une vidéo représente la perception visuelle des objets et/ou locale à un moment donné, c'est-à-dire n/Frate. Chaque pixel (P) dans un plan de vidéo peut être décrit en fonction de l'indice de plan (n) a l'emplacement (x et y) et de l'intensité (r) tel que P (fn, x, y, r) est l'intensité du pixel aux emplacements x et y appartenant au plan fn. Notez que r est le nombre de bits utilisés pour représenter chaque pixel dans le plan (profondeur de pixel). De plus un plan est défini par une largeur (Nx) et une hauteur (Ny) tel que :$1 \leq x \leq Nx$ et $1 \leq x \leq Ny$.

1.3.2.2 Prise de vue

Une prise de vues ou segment (shot) est un ensemble de plusieurs plans enregistrés en continu (sans interruption) par un seul appareil d'enregistrement. Une prise de vues est également considérée comme la plus petite unité d'information visuelle temporelle, Les plans dans un prix de vue (frame intra-shot) contiennent des informations et des caractéristiques visuelles similaires avec des variations temporelles. Ces variations dans le temps entre les éléments de prise de vues peuvent entraîner des changements petits ou importants en raison de l'action entre les marques de début et d'arrêt. Ces changements ne sont dus au fait qu'une prise de vues captures des objets dans le monde réel et que la Sémantiques, la dynamique et la syntaxe deces objets sont fusionnées pour obtenir des plans de prise de vues, tels que le mouvement de l'objet, le mouvement de la caméra ou le fonctionnement de la caméra. De plus, une prise de vues est supposée comporter des objets rigides ou des objets composés de parties rigides reliées entre elles.

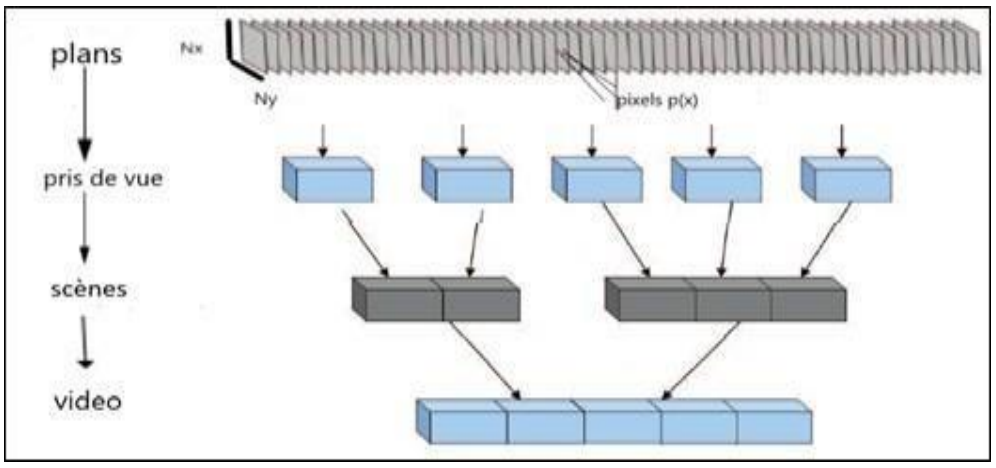


FIGURE 1.2 – Structure hiérarchique d'une vidéo

1.3.2.3 Les scènes

Une scène regroupe plusieurs prises de vue selon un critère sémantique [18]. Les prises de vue et les scènes sont analogues à une phrase et à un paragraphe dans un livre.

1.3.2.4 Les objets

Les objets physiques sont les objets du monde réel qui apparaissent dans les scènes capturées par les caméras. Les objets physiques sont divisés en deux types : les objets de contexte et les objets mobiles [19].

1.4 Les caractéristiques d'un doucement vidéo

Par définition, un document vidéo est défini comme une combinaison de différents flux de données (flux), où visuel et audio sont les deux principaux flux de données. Le flux visuel consiste en une séquence d'images qui sont animés, généralement, en 24 à 30 images par seconde. Le flux audio est un mélange de différents sons (par exemple, mono, stéréo), qui sont généralement échantillonnés de 1600 à 48000 khz. Un troisième flux qui peut être associé est le texte. Il peut apparaître dans certains documents vidéo pour aider les utilisateurs à comprendre la présentation audiovisuelle. Certaines caractéristiques importantes distinguent la vidéo des autres types de données. La vidéo a une résolution plus élevée, un volume de données plus important, un plus grand ensemble de données pouvant être générées, une ambiguïté d'interprétation plus élevée et nécessitent davantage d'efforts d'interprétation. De plus, la vidéo à dimension spatiale et temporelle, alors que le texte n'est que statique non spatial et que l'image est uniquement statique spatiale. De plus, la vidéo n'est pas structurée et contient généralement des relations complexes [21]. Un fichier vidéo est déterminé par les caractéristiques suivantes :

1.4.1 Caractéristiques physiques

ses caractéristiques indique le mode de représentation des données vidéo dans le fichier, une fois le fichier stocké sur un disque dur. Elles peuvent être extraites automatiquement et insérées dans un catalogue de caractéristiques [22] :

1. **Les formats** Un format vidéo décrit comment un dispositif envoie des images d'une vidéo à un autre dispositif, de la même manière qu'un lecteur de DVD envoie des images à un téléviseur, ou un ordinateur à son moniteur. Plus formellement, le format visuel décrit l'ordre et la structure des images qui créent l'image vidéo. Les formats indique la norme de compression utilisée il existe plusieurs types d'un format vidéo mais les plus populaires sont [20] :
   - **Format MPEG :** Dans la plupart des séquences vidéo, la majorité des scènes sont fixes ou bien changent très peu, c'est ce qui s'appelle la redondance temporelle. Lorsque seules les lèvres de l'acteur bougent, presque seuls les pixels de la bouche vont être modifiés d'une image à l'autre, il suffit donc de ne décrire seulement le changement d'une image à l'autre. C'est là la différence majeure entre le MPEG (Moving Pictures Experts Group) et le M-JPEG. Cependant cette méthode aura beaucoup moins d'impact sur une scène d'action. Le groupe MPEG a été établien 1988 dans le but de développer des standards internationaux de compression, décompression, traitement et codage d'images animées et de données audio [20]. Il existe plusieurs standards MPEG [23].
   - **MPEG-1** : développé en 1988, est un standard pour la compression des données vidéo et des canaux audio associés (jusqu'à 2 canaux pour une écoute stéréo). Il permet le stockage de vidéos à un débit de 1.5Mbps dans une qualité proche des cassettes VHS sur un support CD appelé VCD (Vidéo CD).

- **MPEG-2** : un standard dédié originalement à la télévision numérique (HDTV) offrant une qualité élevée à un débit pouvant aller jusqu'à 40 Mbps, et 5 canaux audio. Le MPEG-2 permet de plus une identification et une protection contre le piratage.Il s'agit du format utilisé par les DVD vidéo.
- **MPEG-4** : un standard destiné à permettre le codage de données multimédia sous formes d'objets numériques, afin d'obtenir une plus grande interactivité, ce qui rend son usage particulièrement adapté au Web et aux périphériques mobiles.
- **MPEG-7** : un standard visant à fournir une représentation standard des données audio et visuelles afin de rendre possible la recherche d'information dans de tels flux de données. Ce standard est ainsi également intitulé l'interface de description de contenu multimédia.
- **MPEG-21** : en cours d'élaboration, dont le but est de fournir un cadre de travail (Framework) pour l'ensemble des acteurs du numériques (producteurs, consommateurs, ...) afin de standardiser la gestion de ces contenus, les droits d'accès, les droits d'auteurs. . .
- **Format DivX** : Le format DivX(Digital Video Express) est un format de compression/décompression vidéo permettant d'obtenir des vidéos compressées très peu volumineuses avec une perte de qualité très raisonnable. Ainsi le format DivX permet de stocker un film complet sur un CD-ROM de 650 ou 700 Mo. [20]

2. **L'audio** introduit une autre dimension qui est le temps. En effet par nature (le son est la perception d'une vibration) le son a une durée. La perception de l'information véhiculée par les documents sonores nécessite un certain temps. Pour les documents numériques sonores, la notion de débit devient importante. En effet, pour pouvoir rejouer une musique ou un dialogue il faut disposer d'un certain débit de données. En général ce débit est exprimé en kbit/secondes. Ce débit peut être compressé pour arriver jusqu'à 128 kbit/s [23].

3. **Les paramètres clés**
   - La taille d'un fichier vidéo : est sa dimension sur le disque dur.
   - Le débit :(en images par seconde) est le nombre d'images envoyées ou reçues par unité de temps. Dans notre cas, l'unité de temps est le second. Exigé en général 25 ou 30 images par seconde
   - La définition de l'image : représente la qualité de l'image vidéo liée au nombre de pixels affichés sur une image. Le pixel est la mesure du plus petit détail qui peut être distingué sur un écran
   - La durée d'un fichier vidéo : représente la période de temps nécessaire pour jouer entièrement la vidéo, à son débit normal et sans interruption (pauses ou sautes).
   - Autres informations : le nom d'un fichier vidéo et la date de création sont aussi des caractéristiques physiques d'un fichier vidéo qui peuvent s'avérer utiles dans un processus de recherche [23].

Le tableau suivant représente un exemple de clé physique d'un fichier vidéo :

Nom	Type de compression	Taille	Débit	Format	Définition d'mage	Durée	Date de création
Exemple	MPEG	3.7 Mo	1200 Images par second	.mp4	360×280	2700 seconds	21-06-2021

TABLE 1.1 – exemple d'un catalogue contenant des clés physiques d'un fichier vidéo.

1.4.2 Caractéristiques visuelles

La Caractéristiques visuelles d'un fichier vidéo liées aux images composants une vidéo comme la texture, la forme, le mouvement, le conteur, la couleur [23]

1.4.2.1 La couleur

La couleur est « l'impression produite sur la rétine par un rayon lumineux, qui dépend des longueurs d'onde – de 390 nm (violet) à 760 nm (rouge) – des radiations électromagnétiques qui composent ce rayon ». Pour la représentation des couleurs sur un ordinateur il y a plusieurs modèles de couleurs disponibles. Un modèle de couleurs est une spécification d'un système des coordonnées de couleurs 3D et d'un sous-ensemble visible dans le système de coordonnées où résident toutes les couleurs d'une gamme particulière. Parmi les espaces de couleurs on trouve [23] :

1. **HSV** L'espace colorimétrique HSV représente la couleur sous la forme d'un triplé : teinte H (Hue), saturation S (Saturation) et luminosité V (Value) [26]. Les composantes de teinte (Hue) et de saturation sont intimement liées à la façon dont l'œil humain perçoitles couleurs. La teinte varie de 0 à 1.0, les couleurs correspondantes varient du rouge au jaune, le vert, le cyan, le bleu et le magenta de sorte qu'il y ait réellement des valeurs rouges à la fois à 0 et à 1.0. La saturation varie de 0 à 1.0, les couleurs correspondantes (Hue) varient de non saturer (nuances de gris) à totalement saturé (pas de composant blanc)

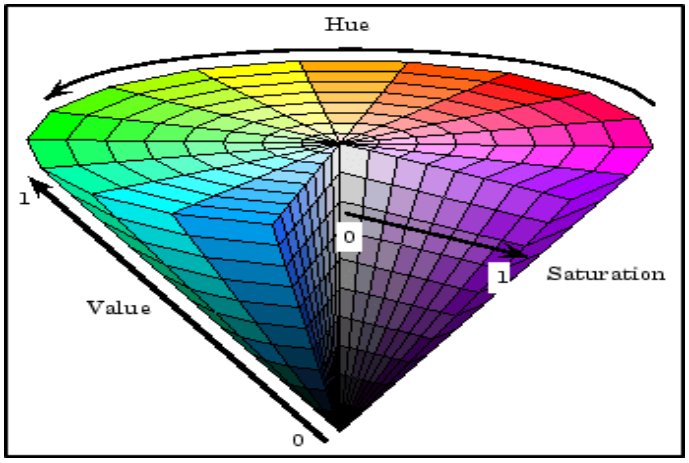


FIGURE 1.3 – Modèle de couleur HSV [26]

2. **RVB** Est l'espace colorimétrique le plus élémentaire pour les vidéos et les images capturées par les caméras. L'espace colorimétrique RVB se compose de trois canaux, Rouge (R), Vert (V) et Bleu (B), qui sont additionnés de différentes manières pour reproduire un large éventail de couleurs. Cet espace colorimétrique est couramment utilisé dans les images numériques. Il est visualisé sous la forme d'un cube 3D où R, V et B sont les trois axes perpendiculaires, comme illustré à la figure suivante [27] :
3. **CIELAB** Le système de couleurs CIELAB, ou CIE L* a* b*, représente la relation quantitative des couleurs sur trois axes : [28]
   — L'axe L* indique la luminosité. 'axea* et b* sont les coordonnées de chromaticité. Sur le diagramme de l'espace colorimétrique :
   — L* est représenté sur un axe vertical avec des valeurs de 0 (noir) à 100 (blanc),
   — L'axe a* comprend a* (positif) et a* (négatif) qui indiquent respectivement les valeurs rouge et verte.

— Les composantes jaune et bleue sont représentées sur l'axe b* par les valeurs b* (Positive) et b* (Négative), respectivement.
— Le centre du plan est neutre ou achromatique. La distance à l'axe central représente la chrominance (Ch), ou saturation de la couleur. L'angle sur les axes de chromaticité représente la teinte (H).

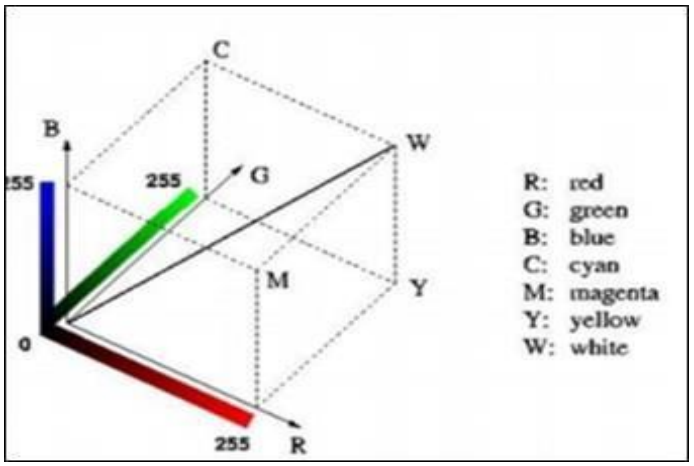


FIGURE 1.4 – Modèle de couleur RGB [27].

— Les valeurs L*, a* et b* peuvent être transcrites en paramètres dermatologiques. La valeur L* est corrélation de niveau de pigmentation de la teinte, et Les valeurs (a*, b*) allant de -120 à +120 pour représenter la chrominance par un angle de teinte et une chromaticité

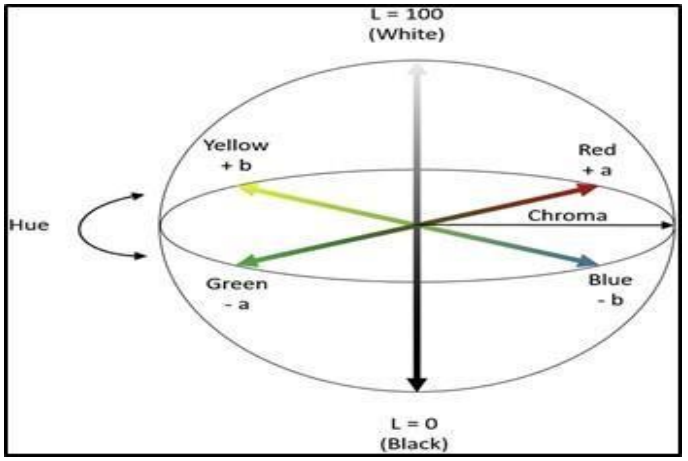


FIGURE 1.5 – Modèle l'espace colorimétrique CIELAB [28].

1.4.2.2 La texture

Est une autre caractéristique visuelle utilisée pour le traitement de la vidéo. Analyser la texture signifie chercher et quantifier les corrélations et relations entre les niveaux de gris des pixels en fonction de la distance qui les sépare. Le traitement des textures se fait par une décomposition de la totalité en plusieurs Élémentaires [25].

1.4.2.3 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, surface ou par des caractéristiques plus précises comme les points de contours. Un descripteur de forme vise à quantifier la forme d'un objet de même que fait la perception humaine [25].

1.4.2.4 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 les plans et/ou 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 de calcul du flux optique.

1.5 L'indexation des vidéos

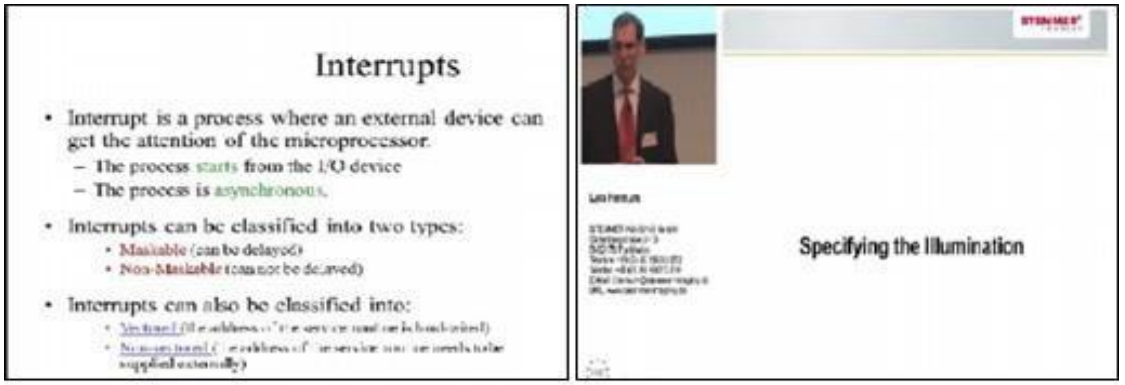
L'indexation est une opération qui décrit et caractérise un document ou un fragment document à l'aide de représentations des concepts contenus dans ce document. Le but de l'indexation est de permettre une recherche efficace et rapide des informations contenues dans un document, les principaux aspects de l'indexation des documents vidéo sont : la représentation et classification [23]. Le choix du contenu ou objet qui sera utilisé pour représenter le contenu sémantique est une tache critique et subjective qui change d'un domaine a autre.

1.5.1 Indexation des vidéos de conférence

Il existe deux types de vidéos de conférences [8] :

1. **Vidéo composée uniquement de diapositives :** ce sont des vidéos composées des diapositives qui contient des informations textuelles ou des images.
2. **Vidéo composée d'une vue de diapositive et d'une vue de l'interlocuteur :** ce sont des vidéos composées des diapositives qui affichent les informations textuelles ou des images en plus d'un interlocuteur qui discute et représente les informations aux audiences.

La figure suivante représente les deux types de vidéo :



(a) Type 1 de vidéo *(b) type 2 de vidéo.*

FIGURE 1.6 – exemple des deux types de vidéos de conférence

Les vidéos de conférence ont deux types de sources d'information, l'audio ou à partir du texte en vidéo. Pour extraire les informations disponibles dans les vidéos de conférence, différentes méthodes sont disponibles [13] :

1. **(ASR) pour la récupération d'informations audio par reconnaissance vocale automatique** : La technique ASR extrait la parole ou la voix de l'audio et le convertit en information textuelle. Informations récupérées l'utilisation de l'ASR dépend de la maîtrise de la parole du locuteur. Il existe différentes indexation et recherche de vidéode conférence méthodes qui utilisent le texte et l'audio extrait de la conférence vidéo.
2. **(OCR) pour les informations textuelles** : la reconnaissance optique de caractères peut être utilisée. La technique ASR extrait la parole ou la voix de l'audio et la convertit en informations textuelles. L'OCR reconnaît le texte des images et produit le texte modifiable. Une vidéo peut être convertie en séquence des images (plans) puis le texte peut être extrait du cadre. Comme la vidéo se compose de plusieurs plans par seconde,la plupart des plans extraites auront un contenu similaire. Par exemple, une vidéo d'une durée de 60 minutes peut comprendre environ des diapositives, mais le nombre des plans présents dans une vidéo peut être de l'ordre de milliers. Il n'est pas nécessaire de traiter tous les cadres pour extraire le texte. L'indexation des vidéos de conférenceest difficile en raison de ses contraintes qui est représenter dans :
   - Les vidéos de conférences contiennent des 'information plus difficile à traiter. Le texte n'est pas séparé du fond. Il est soit superpose (comme les sous-titres) soit inclut dans la scène de l'image. Le fond de l'image peut être très complexe ce qui empêche une séparation facile des caractères.
   - La qualité des vidéos qui peuvent être faible ou élevée en fonction de la résolution de la vidéo. Dans notre cas nous utilisons des vidéos compresser c'est-à-dire une grande perte de la qualité de l'image originale, ce qui signifie une mauvaise résolution. C'est ce qui nous oblige plusieurs méthodes d'amélioration des images.
   - La difficulté de détecter une transition de texte seulement ou cas ou dont la diapositive est en blanc et l'écriture ou le contenue est on noir, cela implique un problèmede la fiabilité de segmentation.
   - Dysfonctionnements externes : panne de courant, changement d'illumination.
   - L'intérêt de l'enseignant seulement par le contenu de vidéo donc il néglige les techniques des montages vidéo et l'utilisation d'équipements simples pour le montage des vidéos. Cela conduit une diminution de qualité.
   - L'activité humaine est très variée : le mouvemente de l'enseignant pendant l'explication et interactions entre les personnes ce qui donne des fausses transitions.

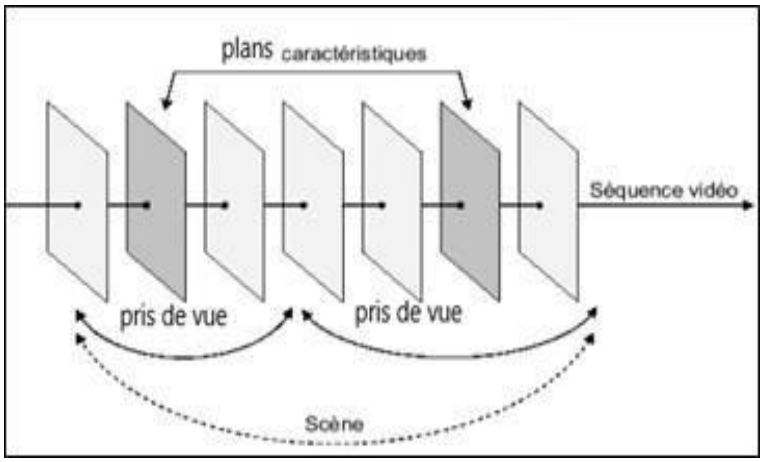


FIGURE 1.7 – Illustration du processus de la segmentation temporelle.

1.5.2 La segmentation temporelle

La segmentation temporelle en plan (En plan) La segmentation en plans est la technique de segmentation temporelle des enregistrements vidéo la plus répandue et la plus utilisée. Les méthodes de détection de changements de plan localisent les images, à travers lesquelles de grandes différences sont observées dans un certain espace de caractéristiques. L'espace de caractéristiques se compose habituellement d'une combinaison de couleur et de mouvement. Les changements de plan peuvent être instantanés (cuts) ou apparaître sur plusieurs images, appelés les effets de transition progressifs, tels que les fondus et les volets. Il est plus facile de détecter des cuts que des effets progressifs. [24].

1.6 Conclusion

Dans ce chapitre nous avons introduit la meilleure source d'apprentissage en ligne qui est les vidéos avec une étude de ses fonctionnalités et de structure ainsi que son indexation. Dansle deuxième chapitre nous présenterons des techniques et méthodes qui permettent de réaliser la segmentation en vue de l'indexation à base de contenu.

CHAPITRE 2

LES SYSTÈMES DE SEGMENTATION TEMPORELLE D'UN DOCUMENT VIDÉO

Sommaire

2.1 Introduction 16
2.2 Les transitions dans les vidéos 16
2.2.1 Définition d'une transition 16
2.2.2 Les types de transition d'une vidéo 16
2.2.3 Transition de diapositive 17
2.3 Architecture générale d'un système de segmentation temporelle . 18
2.3.1 Extraction des caractéristiques 19
2.3.2 Mesures de similarité 19
2.3.3 Détection 19
2.4 Les méthodes de segmentation temporelle des vidéos 19
2.4.1 Amélioration 19
2.4.2 Représentation des informations visuelles 20
2.4.3 Construction du signal de dissemblance/similarité 22
2.4.4 Détection 23
2.5 Métriques d'évaluation d'un système de segmentation temporelle 23
2.6 Conclusion 24

2.1 Introduction

Après une étude approfondie sur la technique vidéo et un aperçu sur l'indexation et la recherche à base du contenu des vidéos, notamment les vidéos de conférence. La détection des points de transition également connue sous le nom de segmentation temporelle des vidéos est considérée comme un processus essentiel permettant l'indexation vidéo. Dans ce chapitre, nous présenterons la notion de transition, l'architecture générale d'un système de segmentation temporelle des vidéos, et nous verrons également les différentes méthodes existantes utilisées pour les différentes étapes.

2.2 Les transitions dans les vidéos

2.2.1 Définition d'une transition

Une vidéo est une concaténation des parties ou ayant une signification mise en œuvre durant le processus de montage vidéo (vidéo editing processus VEP) pour créer une vidéo pendant le processus de production vidéo (vidéo production processus VPP). Les transitions entre les prix de vue sont les moments de passage d'une vue à l'autre. Ils sont ajoutés pendant la postproduction.[17]

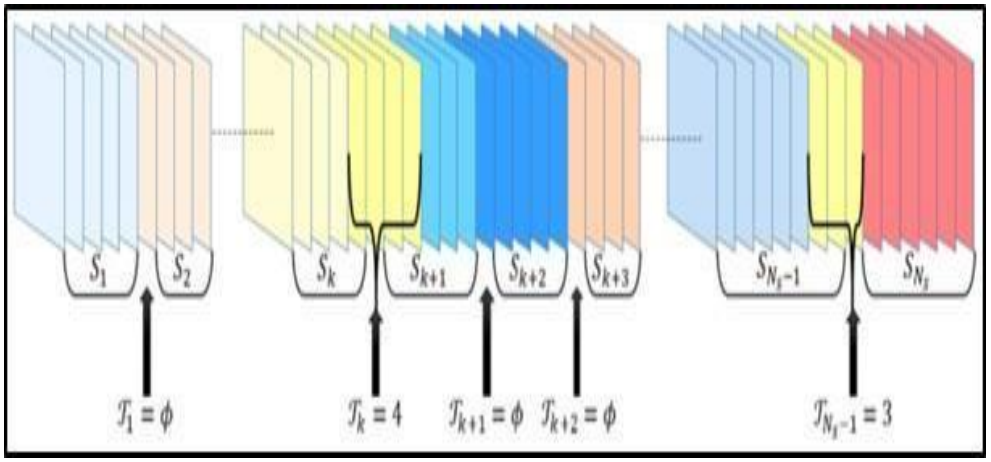


FIGURE 2.1 – Période d'une transition entre les plans [17].

2.2.2 Les types de transition d'une vidéo

Il existe deux types de transition à savoir les transitions dures et douces [17]. La figure suivante résume les différents types de transition d'une vidéo :

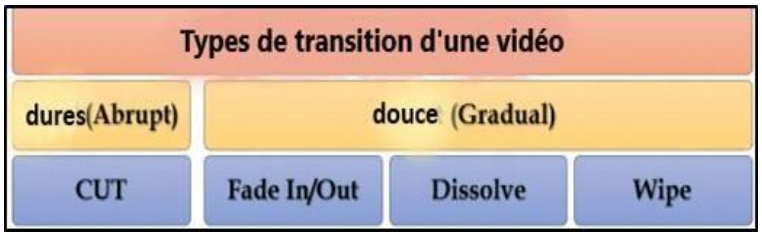


FIGURE 2.2 – Les types de transitions d'une vidéo [17].

2.2.2.1 Transitions dures

Est également connu sous le nom : une transition coupée ou abrupte. C'est le type le plus fréquent qui marque le fait que deux plans successifs sont directement concaténés sans aucun montage (effets spéciaux). Un exemple d'une transition dure est donné sur la figure 2.3.on remarque l'absence de plans intermédiaires entre les prises de vue mais le passage se fait directement.



FIGURE 2.3 – exemple d'une transition dure entre les deux plans 1952,1953.

2.2.2.2 Transitions douces

Également connu sous le nom : une transition graduelle ou continue. La transition graduelle se produit lorsque deux prises de vues sont combinées en utilisant des effets spéciaux, ces effets peuvent inclure un ou plusieurs plans entre deux prises de vues consécutives. La transition graduelle couvre de nombreux types de transition, y compris la transition dissoudre, fondu entrant, fondu sortant et transition d'effacement.[17]

1. Dans la transition dissoudre (dissolve), les valeurs d'intensité des pixels diminuent progressivement (diminuent) à partir d'un prix de vues k, et l'intensité des pixels apparaît progressivement (apparaît) à partir du prix de vue suivante (changement entre les prises de vues sont partiellement visibles).
2. Dans la transition fondue entrante (fade in), les valeurs d'intensité des pixels du prix de vue $S_(k + 1)$ émergent progressivement d'un plan d'intensité fixe. En revanche, le plan précédant SK est directement modifié par le plan d'intensité fixe.
3. Dans la transition fondue sortante (fade out), les valeurs d'intensité des pixels sont progressivement modifiées d'un prix de vue S_K à un plan d'intensité fixe. Le prochain pris de vue $S_(k + 1)$ apparaît instantanément après le plan d'intensité fixe.
4. Dans la transition d'effacement (wipe) les pixels de la prise de vues en cours sont progressivement remplacés par les pixels correspondants de la prise de vues suivante. Par exemple, le remplacement progressif des pixels de colonne de gauche à droite du cadre est considéré comme un effacement typique.

La figure suivante 2.4 montre un exemple d'une transition graduelle de type fondu entrante :



FIGURE 2.4 – exemple de transition fondu entrant avec 21 plans impliqués dans la transition (n=22-1).

2.2.3 Transition de diapositive

La transition diapositive (Diapositive transition) est un type des transitions dures. Les transitions de diapositives se produisent lorsque l'interlocuteur passe d'une diapositive à l'autre afin d'expliquer de son contenu. Ce type de transition est détecté dans les vidéos éducatives (les vidéos de conférences). La figure 2.5représente un exemple d'une transition de diapositive dans les 2 types de vidéos de conférences.

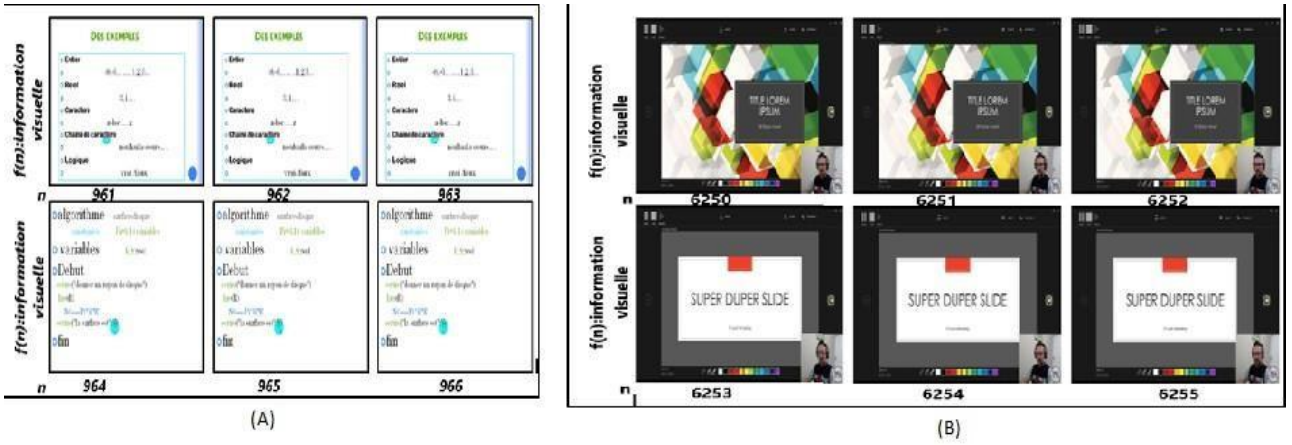


FIGURE 2.5 – exemple de transitions diapositive (A) vidéo de type 1 (B) vidéo de type 2.

2.2 Architecture générale d'un système de segmentation temporelle

Un système de segmentation temporelle des vidéos : est un système qui détecte des transitions (coup limites) dans une séquence vidéo [17]. Le système de segmentation temporelle des vidéos prend comme une entrée une base des vidéos, ces vidéos seront ensuite traitées par des algorithmes de détection des transitions Enfin, ce système va détecter les transitions entre deux prises de vues, une transition peut être classée en deux types : transitions dures (coupée)et douce (graduelle). Généralement, le système de segmentation temporelle comprend trois phases :

(1) Extraction des caractéristiques, (2) Mesure de similarité et (3) Détection. Chaque phase peut inclure des étapes de prétraitement et/ou de post-traitement.[17]

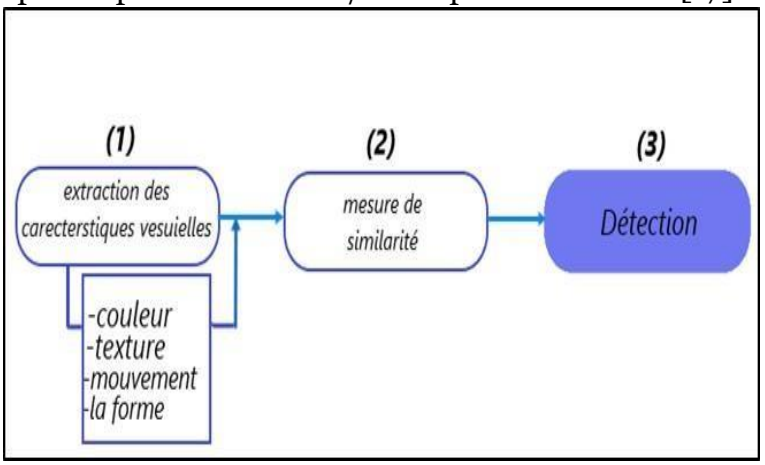


FIGURE 2.6 – architecture générale de système temporelle des vidéos.

2.2.1 Extraction des caractéristiques

La représentation des informations visuelles (Représentation of visual Information ROVI) pour un plan de vidéo est effectuée en extrayant ses caractéristiques visuelles comme la cou- leur, texture, forme et le mouvement. Le résultat est une représentation numérique exploitable par les étapes suivantes. L'objectif est d'utiliser une méthode d'extraction appropriée pour les caractéristiques des vidéos et aux conditions d'acquisitions en reflétant les changements significatifs dans les transitions.

2.2.2 Mesures de similarité

La mesure de similarité permet de calculer la distance (dissemblance /similarité) entre les caractéristiques de deux plans successifs ($p_i et p_{(i}e + 1)$). Il existe plusieurs mesures de distance, qui permettent de calculer la différence des caractéristiques visuelles. Ainsi pour des plans on similaires on obtient une valeur élevée de distance, dans le cas de plans similaires la valeur doit diminuer.

2.2.3 Détection

Il s'agit d'une de classification des plans à base de la distance calculée le signal de dissemblance/similarité est testée avec une valeur de seuil .

2.3 Les méthodes de segmentation temporelle des vidéos

Il existe de nombreuses méthodes de détection de segmentation temporelle, en ce qui suit nous présenterons quelques-unes pour chaque étape du système.

2.3.1 Amélioration

L'amélioration c'est une méthode n'appartient pas ou système segmentation temporelle, mais elle est considérée comme une étape très importante dans le prétraitement des plans ce qui améliore les performances du système. L'amélioration consiste à ajuster un ou plusieurs paramètres des plans pour modifier ses caractéristiques pour avoir autres de haute qualité.

2.3.1.1 La suppression de la brume

Une méthode d'amélioration d'image basée sur la suppression de la brume, un ensemble d'images détaillées et sous-exposées dérivées d'une seule image floue est utilisée comme entrées pour la fusion d'images. Les cartes de poids correspondants sont calculées selon le canal sombre avant (dark channel prior DCP), qui est bien reconnue comme un bon indicateur de brume.Ensuite, le processus de fusion est simplement équivalent à une somme pondérée d'images et de cartes de poids. Enfin, une méthode de post-traitement connue sous le nom de remappage adaptatif des tonalités est utilisée pour étendre la plage dynamique. Ainsi, l'algorithme proposé est efficace du point de vue informatique et consciente de la brume, tandis que son homologue matériel compact est capable de gérer des vidéos en temps réel [35].

2.3.1.2 Le double éclairage

Cette méthode basée sur un L'algorithme dédouble éclairage qui est basé sur les méthodes de RETINEX. Ou les méthodes de RETINEX Effectues une transformation spectrale non linéaire qui synthétise une forte amélioration du contraste local et une stabilité des couleurs [37]. Et l'algorithme de double éclairage était donné l'image convertie vers l'espace CILAB et l'inverser, d'abord il effectuée l'estimation d'éclairage double pour obtenir les éclairages avant et arrière, à partir de laquelle on récupère Image corrigée de sous-exposition et surexposition en- suite, une fusion d'images multi-exposition efficace est appliquée pour mélanger de manière transparente les parties visuellement les mieux exposées dans les deux images de correction et l'image d'entrer [38].

2.3.2 Représentation des informations visuelles

2.3.2.1 Méthode basée sur les pixels

La méthode basée sur les pixels (PBA) ou la comparaison au niveau des pixels est utilisée comme ROVI directement à partir des intensités de pixels des plans vidéo. Le PBA consiste à calculer la différence entre deux pixels (aux emplacements x et y) de deux plans vidéo consécutifs ((*pletp(le* + 1)). Dans l'étape suivante de PBA, la somme totale des différences de pixels est déterminée et comparée à un seuil. Une transition est déclarée si la la somme dépasse leseuil sélectionné. Généralement, cette méthode est basée sur l'intensité des pixels, et elle est déterminée par l'équation (2.1) suivante [17] :

1

$$\boldsymbol{D(p_l, p_{l+1}) = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |p(p_l, x, y) - p(p_{l+1}, x, y)|} \qquad (2.1)$$

Tel que :

1. $\boldsymbol{(p_l, p_(l+1))}$: les deux plans successifs.
2. $\boldsymbol{N(x)N^JY}$: la largeur et la hauteur de plan, respectivement.
3. $\boldsymbol{P(.)}$: L'intensité des pixels.

Pour détecter une transition, en comparant la somme des différences absolues du pixel total (équations (2.1)) avec une valeur seuil. Lorsque la somme était supérieure au seuil, la transition était déclarée ; dans le cas contraire, un plan de pris de vue a été pris en compte Parmi les méthodes basées sur les pixels il existe la méthode de déférence des pixels globaux : dans cette méthode, les plans sont extraits de la vidéo à différents intervalles de temps et sont convertis au binaire, puis la différence au niveau du pixel est calculée par la méthode de

Soustraction de plan. Le nombre de pixels non nuls de plan résultant est normalisé avec la taille de plan[13].

$$\boldsymbol{D_I = \sum_{x=0}^{w} \sum_{y=0}^{h} |b_i(x,y) - b_{i-1}(x,y)|} \qquad (2.2)$$

$$\boldsymbol{p_i = \frac{D_i}{w \times h}} \qquad (2.3)$$

1. $b_i(x,y)$
2. w×h : Est la taille de l'image.
3. p_i : La valeur qui tester avec le seuil.

2.3.2.2 Méthode basée sur l'histogramme

L'approche basée sur l'histogramme (histogram-based approach HBA) utilise le niveau de gris pour la caractérisation des plans et la détection des transitions. Dans cette approche, les histogrammes de niveaux de gris des plans successifs sont calculés et comparés à l'aide d'une métrique de différence d'histogramme pour détecter les deux types de transition coupée et graduelle via des seuils bas et hauts (SB et SH respectivement). La transition graduelle est détectée lorsque HDM est supérieur à SB. [17,13]

$$\boldsymbol{HDM(p_l, p_{l+1}) = \sum_{j=1}^{Nh} |H(p_l, j) - H(p_{l+1}, j)|} \qquad (2.4)$$

N_h: Est le nombre total de niveaux de gris possibles

$\boldsymbol{H(p_l, j)}$: Est la valeur de l'histogramme pour le niveau de gris j au plan n.

la méthode inclue les étapes suivants :

1. Sélection une vidéo en entrée.
2. Extraction les plans à différents intervalles de temps.
3. le calcul de la différence d'histogramme de couleur entre chaque paire des plans adjacents extraits à l'aide de l'équation (2.2)
4. Si la différence d'histogramme $HDM(p_l, p_{l+1}) > valeur$ du seuil Ensuite, la transition est détectée.
5. Répétions des étapes précédentes jusqu'à la fin de la vidéo.

2.3.2.3 Méthodes basées sur les conteurs

La méthode basée sur les conteurs (Edge-bases approchent EBA) considère une caractéristique de bas niveau d'un plan. Cette méthode est mise en œuvre pour détecter une transition coupée et graduelle. Dans ces méthodes, la transition est déclarée lorsque les emplacements des bords du plan courant présentent une grande différence avec les bords du plan précédent qui ont disparu. Détection des contours (y compris les contours nouveaux et précédents),), rap- port de changement de contour (Edge change ratio ECR) et mouvement compensation sont les processus requis pour calculer les changements de bord [17]. Il existe autres méthodes de segmentation temporelle des vidéos telles que : méthode basée sur la texture, méthode basée sur le mouvement.

2.3.3 Construction du signal de dissemblance/similarité

2.3.3.1 La formule de différence de couleur CIEDE 2000

La formule de différence de couleur CIEDE2000 basée sur l'espace colorimétrique CIELAB. Il comprend non seulement des fonctions de pondération de luminosité, de chrominance et de teinte, mais également un terme interactif entre les différences de chrominance et de teinte pour améliorer les performances des couleurs bleues et les et les performances des couleurs grises la formule CIED2000 était testée avec les autres techniques avancés basés sur CIELAB. Il a surperformé CMC et CIE94. [32] la formule de différence de couleur CIEDE2000 est basée sur l'espace colorimétrique CIELAB. Étant donné une paire de valeurs de couleur dans l'espace CIELAB $l^*_{1,}a^*_{1,}b^*_{1;}l^*_2$ et l^*_2, a^*_2, b^*_2 On note le la distance CIEDE2000 différence entre les Deux couleurs comme suit [31] :

$$\Delta E00 = \Delta E00\ (l^*_{1,}a^*_{1,}b^*_{1;}l^*_2\ ; l^*_2, a^*_2, b^*_2) \quad (2.5)$$

Le diagramme suivant résume les étapes de CIEDE2000 :

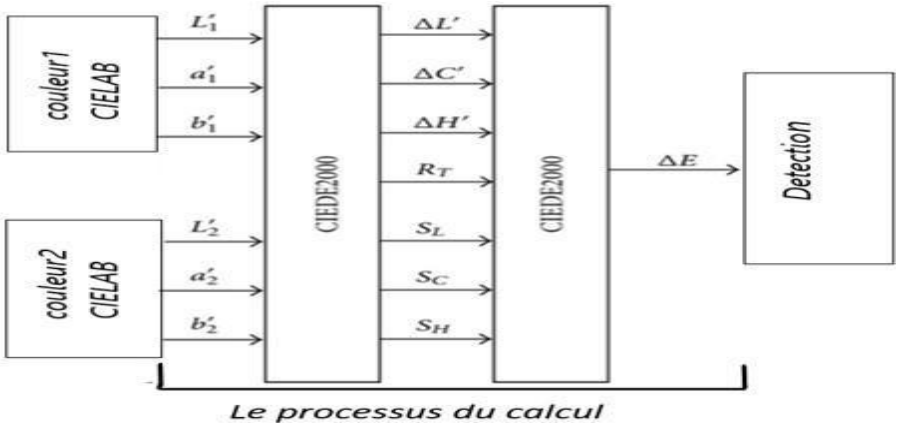


FIGURE 2.7 – Les étape de calcul du CIEDE2000.

2.3.3.2 Les avantages du CIEDE2000

Les avantages de CIEDE2000 sont [35] :

— Actuellement, CIEDE2000 est considéré comme le meilleur modèle de différence de cou- leur uniforme coïncidant avec la perception visuelle subjective. Il normalise la luminosité,la teinte et la saturation de la perception visuelle à la même unité.

— L'espace de différence de couleur uniforme a été proposé pour établir une relation linéaire entre les changements de luminosité, de teinte et de saturation et la perception visuelleà travers l'expérience de perception visuelle subjective. Il a été démontré expérimentalement que les résultats d'évaluation de l'espace de différence de couleur uniforme sont bien meilleurs que l'espace de couleur de l'adversaire et l'espace de couleur YUV.

— Avec l'espace CIELAB, nous pouvons obtenir directement un résultat numérique, nommé paramètre de différence de couleur E, qui peut refléter la différence de couleur entre deux images. De nombreux chercheurs ont exploité ce paramètre de différence de couleur E de

la formule CIEDE2000 dans l'évaluation de la qualité d'image.

— CIED2000 donne une métrique perceptuelle ment uniforme afin que le seuil soit constant sur toute la gamme (gamme de couleurs). Sinon, le seuil sera fonction de la couleur de référence, encombrant comme guide pratique.

2.3.4 Détection

2.3.4.1 Estimation du seuil

Le seuil est la valeur minimale du stimulus à partir de laquelle un individu parvient à détecter une stimulation. Du fait de la variabilité du rapport signal sur bruit consister à tester pour chaque pixel de plan si sa valeur est supérieure ou inférieure à un certain seuil. Il est aussi possible d'utiliser plusieurs seuils pour identifier les limites des plans en fonction des similitudes d'images, un seuil doit être établi. Il peut être global ou adaptatif : [30]

1. **Seuil global** c'est même seuil prédéfini sur une vidéo entière, où la valeur est généralement définie de manière empirique. La principale limitation de l'utilisation d'un seuil fixe est que la variation du contenu local n'est pas correctement reflétée dans la détection des limites.
2. **Adaptatif** est déterminée par le calcul de la valeur de seuil pour chaque région fractionnaire de l'image, de sorte que chaque région fractionnaire a une valeur de seuil différente.

2.3.4.2 Classifications

Classification permet de classifier les transitions et non transitions. Les transitions détectées sont enregistrées et classifiées selon le type de transition : graduelle, coupée, diapositive, et par le numéro du plan dans laquelle la transition s'est produite.

2.4 Métriques d'évaluation d'un système de segmentation temporelle

Les performances d'un système de segmentation temporelle sont mesurées à l'aide des trois métriques : rappel, précision et F-mesure qui sont calculées avec les formules ci-dessous [31] :

1. La précision (ou valeur prédictive positive) est la proportion des items pertinents parmi l'ensemble des items proposés :

$$Précision = N_C/(N_C + N_M) \tag{2.6}$$

2. Le rappel (ou sensibilité) est la proportion des items pertinents proposés parmi l'ensemble des items pertinents.

$$Rappel = N_C/(N_C + N_F) \tag{2.7}$$

Tel que :

N_C: Les vrais transitions détectées par le système.

N_M: Les transitions nondétectées par le système.

N_F :Les fausse transitions détectées par le système.

3. F1-mesure : Une mesure qui combine la précision et le rappel est leur moyenne harmonique :

$$F1_mesure = (2 * Précision * Rappel)/(Précision + Rappel) \tag{2.8}$$

2.5 Conclusion

Dans ce chapitre, nous avons présenté les différents types de transition dans une vidéo et les méthodes de segmentation temporelle des vidéos, les différents espaces de couleur présentent des caractéristiques utiles pour la segmentation, en ce qui suit l'espace CIELAB est utilisé avec la formule CIEDE2000. Dans le chapitre suivant nous présenterons notre système de détection des points de transition.

CHAPITRE 3

CONCEPTION

Sommaire

3.1 Introduction 26
3.2 Architecture générale du système 26
3.3 Description détaillée 27
3.3.1 Phase 1 : le prétraitement 27
3.3.2 Phase 2 : L'analyse 29
3.3.3 Phase 3 : La détection des points de transition 31
3.4 Conclusion 32

3.1 Introduction

Dans les deux premiers chapitres, on a étudiée la structure des vidéos, la notion de transition ainsi que les différentes méthodes utilisées pour la détection. L'étude détaillée nous a permisde choisir parmi les méthodes existantes les méthodes plus efficaces permettent la détection des points de transition et les adaptées aux vidéos à usages éducatives. Dans ce chapitre, nous allons présenter l'architecture de notre système ainsi que ces différentes étapes.

3.2 Architecture générale du système

On propose un système de détection de transitions dans les vidéos de conférences dont l'objectif est qu'à partir d'une vidéo, extraire un ensemble de points de transition représentés sous forme de plans vidéos. Notre système, est composé de trois phases, commençant par le prétraitement pour améliorer la qualité des plans, puis l'analyse qui comprend l'extraction des caractéristiques et calcul de la distance. Et enfin la détection qui consiste à vérifier le point de transition par rapport au seuil. L'architecture générale du système est donnée par la figure suivante 3.1

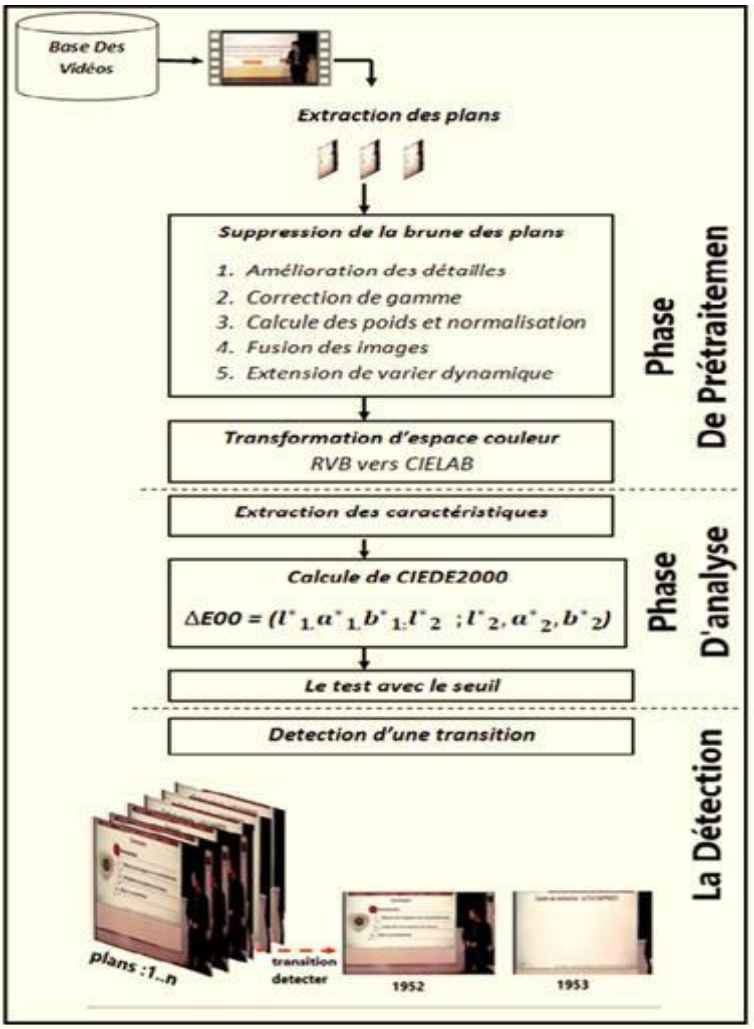


FIGURE 3.1 – Architecture générale du système

3.3 Description détaillée

3.3.1 Phase 1 : le prétraitement

Les plans vidéo présentent plusieurs dégradations relatives aux conditions d'acquisition et aux mouvements, ce qui impose l'utilisation des méthodes de prétraitement pour la préparation des plans à la phase d'analyse. La phase de prétraitement comporte deux étapes : l'amélioration de la qualité des plans et la transformation d'espace de couleurs.

3.3.1.1 Suppression de la brume des plans

Les vidéos de conférence sont de mauvaise qualité et subissent un manque de lumière, ainsi, il est préférable d'améliorer les plans avant de passer aux autres traitements pour garantir des meilleurs résultats. Dans notre système on utilise la méthode d'élimination de brume [35] exploite principalement les techniques usuelles de traitement d'images pour permettre l'amélioration des détails, la fusion d'images à exposition multiples et le remappage des plans. La figure 4 présente les étapes de cette méthode.

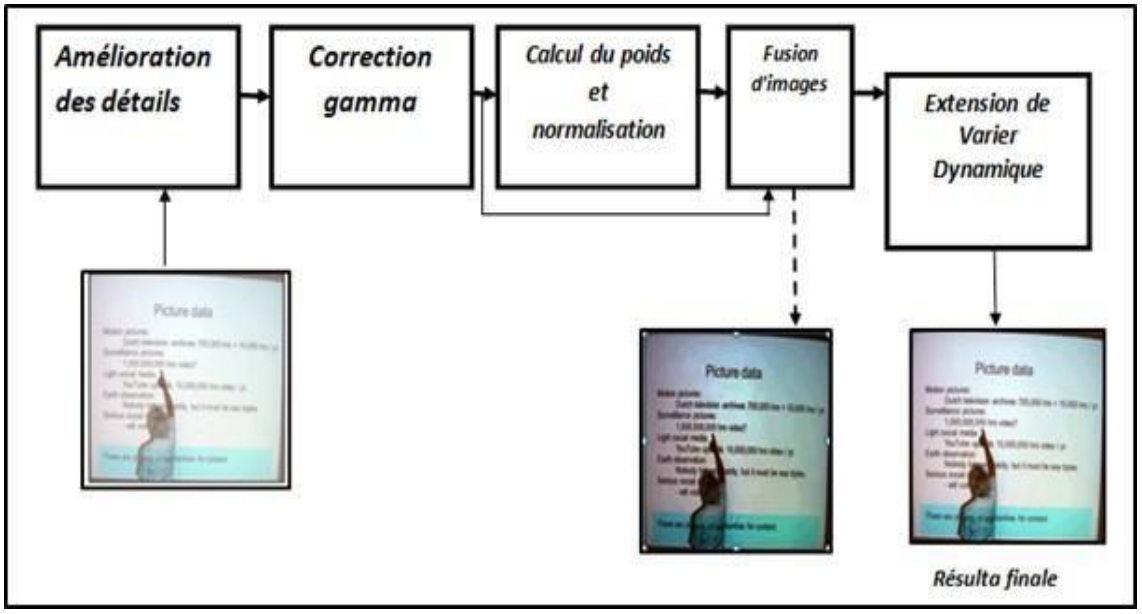


FIGURE 3.2 – Description générale de la méthode de suppression de brume. La

méthode de suppression de brume [35 ,36] est constituer de plusieurs étapes :

1. **L'amélioration des détails** : également appelée amélioration de la netteté, est une technique de traitement d'images qui implique trois étapes principales :
   — La décomposition de l'image d'entrée en signaux d'arrière-plan et de détail.
   — Multiplier ce dernier (c'est-à-dire le signal de détail) par un facteur adéquat
   — Rajouter les détails améliorés au signal de fond. Premièrement L'image d'entrée est convertie dans l'espace colorimétrique YCbCr pour améliorer le canal de luminance. La raison en est que ce canal contient beaucoup plus d'informations de bord haute fréquence que celles de chaque canal R, V et B. de plus, le format YCbCr réduit les ressources de calcul liés au mouvement des données de l'algorithme. Par la suite, les données d'image sont reconverties dans l'espace colorimétrique RVB pour être récupérées dans le module de correction gamma.

2. **Correction gamma** : est une technique pour sous-exposer artificiellement la lumière dans les images. Il s'agit d'une opération de traitement d'image non linéaire qui est généralement définie par l'expression de loi suivante :

$$I_u^c(x) = [I^c]^\gamma \quad (3.1)$$

Où, l'exposant **c** désigne un canal de couleur de l'image d'entrée (c'est-à-dire, **c**∈ **R, V, B)**, l'indice **u** représente la sous-exposition et est une constante représentant le degré d'exposition. Étant donné les données d'image normalisées de la varier [0, 1] la surexposition et la sous-exposition sont représentées par $\gamma < 1 et \gamma > 1$respectivement.

3. **Calcul du poids et normalisation** La sélection d'un schéma de pondération approprié dans la fusion d'images, dépend de l'objectif de l'algorithme de suppression de la brume d'image unique, ce dernier a fusionné une séquence d'images multi-expositions en une image de haute qualité. Le canal sombre constitue une base solide pour dériver un schéma de pondération en tenant compte de la brume. En effet, la clarté et la couleur des scènes capturées conduisent à l'existence de pixels sombres dont l'intensité est proche de zéro sur au moins un canal de couleur. Le canal sombre est défini comme suit :

$$I^{DCP}(x) = min_{(y \in \Omega(x))} min_{(c \in R,G,B)} [I^c]^\gamma \quad (3.2)$$

Où, **y** représentent les coordonnées des pixels dans la fenêtre carrée Ω(x) centrée en x. Cette équation permet aux pixels sans brume d'avoir un DCP (Dark Channel Prior) canal sombre prédéfini extrêmement faible, tandis que les pixels brumeux présentent un DCP important en raison des effets de brume.

4. **La fusion d'images** : La fusion d'images est réalisée à une seule échelle comme une simple somme pondérée d'images sous-exposées. La fusion est donnée par l'équation Suivante :

$$I^{DCP}(x) = min_{y \in \Omega(x)} \{ min_{c \in \{R,G,B\}} \} [I^c]^\gamma \quad (3.3)$$

Où, J est l'image floutée, I_u^iest l'une des images sous-exposées,$w_i$est l'une des cartes de poids correspondantes. Une interprétation détaillée est ensuite effectuée a fin de prendre en charge l'utilisation de la fusion d'images à une seule échelle. En supposant qu'il existe deux images sous-exposées $\{I_1^1, I_2^2\}$ dérivées de l'image d'entrée unique, deux cartes de poids correspondantes $\{w_1, w_2\}$ sont calculées sur la base du canal sombre avant

5. **Extension de Varier Dynamique :** Le ciel assombri après la fusion d'images nécessite une amélioration de la luminance ; en outre, l'ensemble de l'image doit être amélioré. Bien que l'utilisation d'images sous-exposées soit bénéfique pour l'élimination du voile, toutefois, elle entraîne un effet secondaire indésirable à savoir un assombrissement considérable de l'image entière.

Enfin, on obtient le résultat final de l'algorithme de suppression de la brume, un exemple de résultat est présenté sur la figure 3.3.

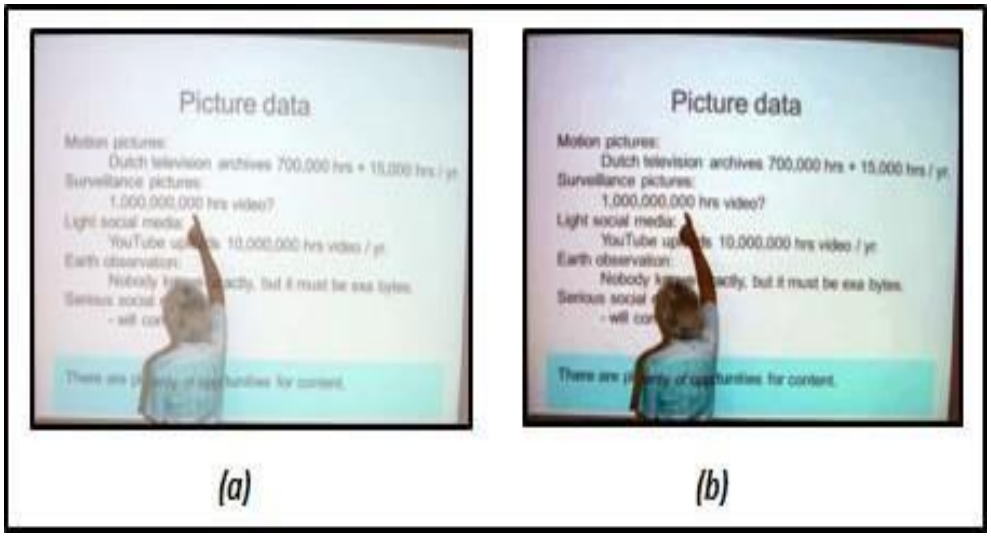


FIGURE 3.3 – Exemples de résultat de suppression du brume(A) plan original, (B)plan après la suppression de la brume.

3.3.1.2 Transformation d'espace de couleurs

Dans cette étape, on effectue une transformation d'espaces colorimétriques BVR vers l'espace colorimétrique CIELAB pour un calcul stable et efficace. En effet, l'espace CIELAB présente plusieurs avantages : il permet d'approximer toutes les couleurs disponibles perçues par l'œil humain, en plus, il permet de séparer les composants de couleur de celle de l'intensité. Cette séparation garantit une robustesse aux changements d'éclairage et la suppression des ombres [31]. La figure suivante montre un exemple de conversion d'espace de couleur de BVR vers CIELAB.

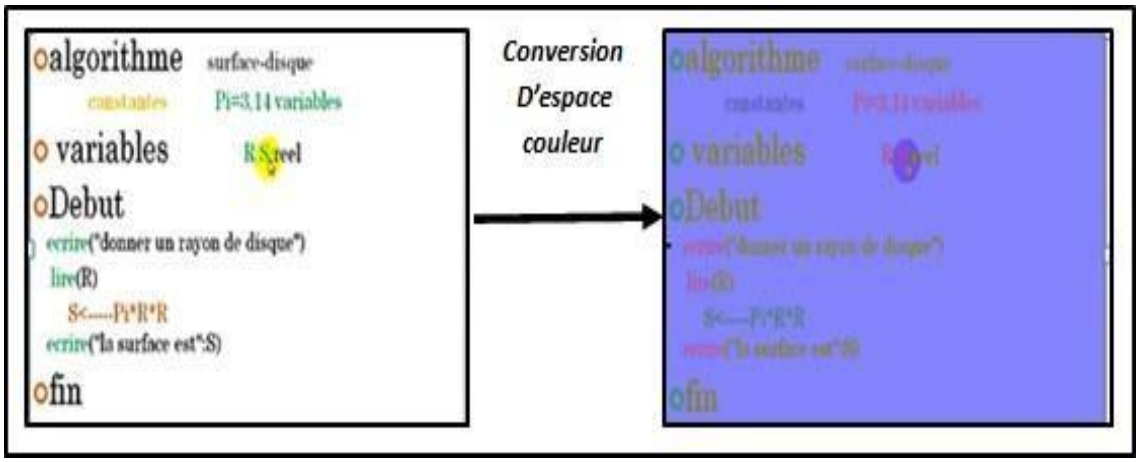


FIGURE 3.4 – Transformation d'espace de couleurs, (a) image dans l'espace BVR(b) image dans l'espace CILAB.

3.3.2 Phase 2 : L'analyse

Cette phase consiste à extraire les caractéristiques visuelles des plans pour comparer leurs contenus via la formule de CIEDE2000 en utilisant seuil. La phase d'analyse inclue les étapes suivantes : extraction des caractéristiques et Calcul de la distanceCIEDE2000.

3.3.2.1 Extraction des caractéristiques

L'extraction des caractéristiques est une étape critique dans le processus de détection, elle est effectuée en extrayant un descripteur à base d'informations colorimétriques pour chaque plan de vidéo afin de caractériser leurs contenues.

Cette étape passe par deux pas à savoir :

1. On devise les plans ont quatre sous-plans de taille identique, chaque sous plans a un certain nombre de pixels, et chaque pixel contient trois canaux de couleurs L, A ou B.
2. Pour chaque sous plan (1, 2, 3,4) on calcule la moyenne arithmétique des canaux L,A et B en utilisant l'équation suivante :

$$MC = \frac{1}{M \times N}\left(\sum_{m=1}^{M}\sum_{n=1}^{N} I(m,n)\right) \qquad (3.4)$$

Où :I(m,n)est la valeur du pixel dans chaque sous plan.

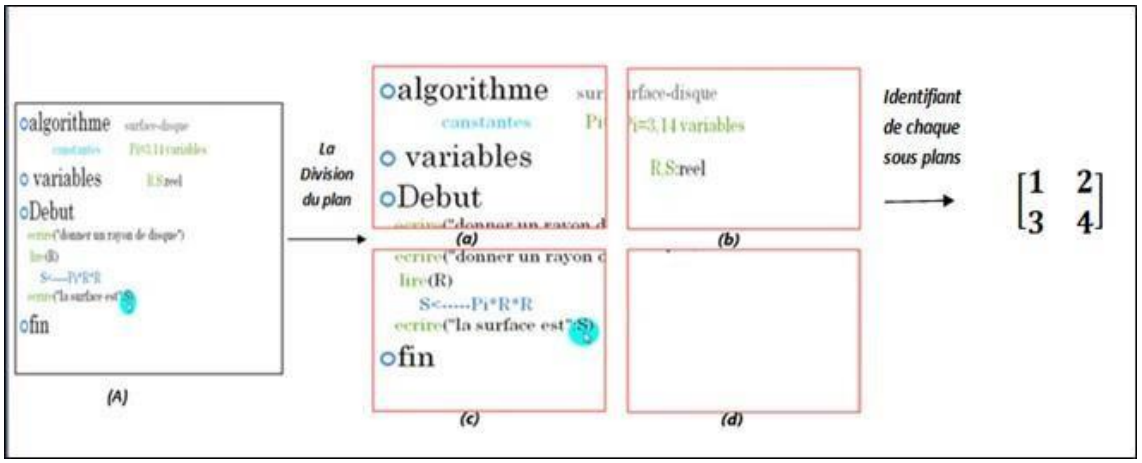


FIGURE 3.5 – Exemple de la division des plans : (A) plan original, (a) sous plan 1, (b) sous plan 2, (c)sous plan 3 et (d) sous plan 4.

3.3.2.2 Calcul de la distance CIEDE2000

La détection des points de transition se base sur la détection d'une dissimilarité entre deux plans consécutifs. La formule CIEDE 2000 est utilisée pour calculer une valeur $\Delta E_{finalle}$ qui représente la différence de couleur entre deux plans dans l'espace colorimétrique CIELAB, cette valeur sera utilisée pour la détection dans la phase suivante. Etant donné une paire de valeurs de couleurs dans l'espace CIELAB, la différence CIEDE2000 est donnée par l'équation (3.6)

$$\Delta E00 = \Delta E00\ (l^*_1, a^*_1, b^*_1; l^*_2\ ; l^*_2, a^*_2, b^*_2) \quad (3.5)$$

$$\Delta E = \sqrt{\left(\frac{\Delta l'}{K_l S_l}\right)^2 + \left(\frac{\Delta c'}{K_c S_c}\right)^2 + \left(\frac{\Delta h'}{K_h S_h}\right)^2 + R_{t_2}\left(\frac{\Delta c'}{K_c S_c}\right)^2 + \left(\frac{\Delta h'}{K_h S_h}\right)^2} \quad (3.6)$$

Tel que :

— **Δl',Δc' et Δh'** sont les différences de luminosité, de saturation et de teinte respectivement, pour une paire d'échantillons dans CIEDE2000. Ces derniers sont donnés par les formules suivantes respectivement :

$$\Delta L' = L'_b - L'_s \tag{3.7}$$

$$\Delta C' = C'_b - C'_s \tag{3.8}$$

$$\Delta H' = 2\sqrt{C'_b C_s}\, sin\left(\frac{\Delta h'}{2}\right) \tag{3.9}$$

— R_t est une fonction de rotation qui donne l'interaction entre la teinte et la saturation dans la zone bleue elle est donnée par la formule suivante :

$$R_t = -sin(2\Delta\sigma)R_c \tag{3.10}$$

— La luminosité (S_l) est définie par la formule :

— La chrominance (S_c)est définie par la formule :

$$S_c = 1 + 0.045(C^J) \tag{3.11}$$

— La teinte (S_h)est définie par la formule :

$$S_H = 1 + 0.01C^J T \tag{3.12}$$

(S_l), (S_c), Et (S_h) sont utilisées pour définir la compensation de la luminosité, la chrominance et de la teinte respectivement, D'autre part les autres paramètres de pondération manuelle sont désignés par : K_l, K_C et K_H. Où : $K_l = 1$, $K_C = 1 et K_h = 1$.

Dans notre cas la formule CIEDE2000 permet de calculer la différence entre les quatre sous plans de deux plans consécutifs.

— ΔE_1 : La différence entre (sous plan 1) du (plan i) et (sous plan 1) de (plan i+1)

— ΔE_2 : La différence entre (sous plan 2) du (plan i) et (sous plan 2) de (plan i+1)

— ΔE_3 : La différence entre (sous plan 3) du (plan i) et (sous plan 3) de (plan i+1)

— ΔE_4 : La différence entre (sous plan 4) du (plan i) et (sous plan 4) de (plan i+1) Par l'équation suivante :

$$\Delta E_i = \sqrt{\left(\frac{\Delta l'}{K_l S_l}\right)^2 + \left(\frac{\Delta c'}{K_c S_c}\right)^2 + \left(\frac{\Delta h'}{K_h S_h}\right)^2 + R_t\left(\frac{\Delta c'}{K_c S_c}\right)^2 + \left(\frac{\Delta h'}{K_h S_h}\right)^2} \tag{3.13}$$

Enfin on calcule la moyenne arithmétique des différences de quatre sous-plans par l'équation suivante :

$$\Delta E_{finalle} = (\Delta E_1 + \Delta E_2 + \Delta E_3 + \Delta E_4)/4 \quad (3.14)$$

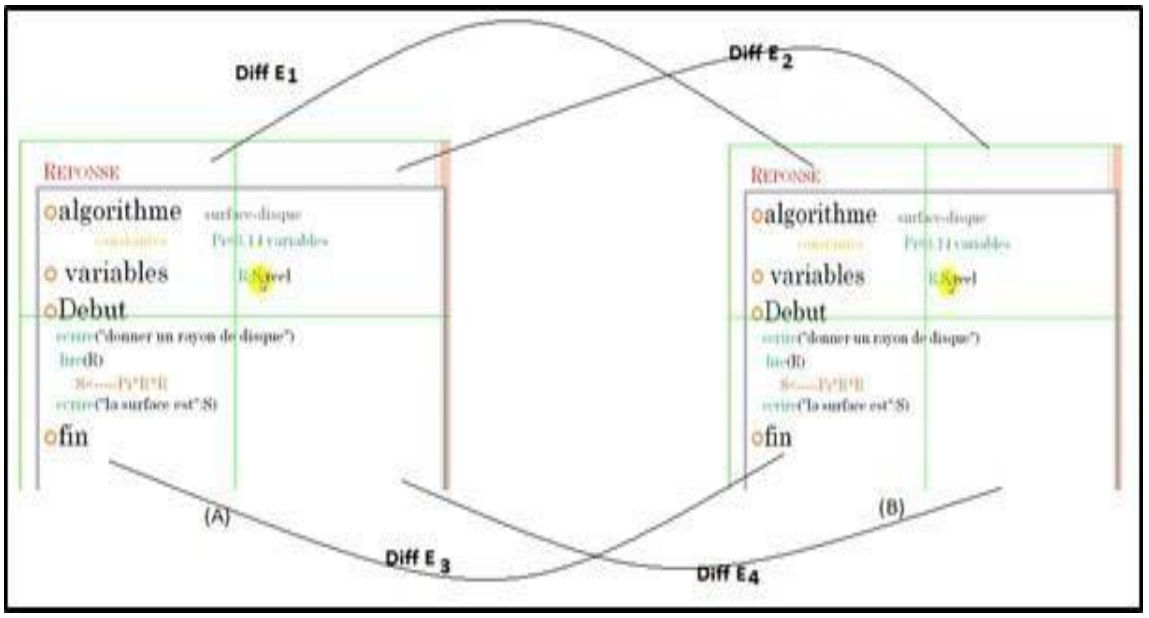


FIGURE 3.6 – le processus de CIEDE2000 entre plan (A) et le plan (B).

3.3.3 Phase 3 : La détection des points de transition

La valeur de $\Delta E_{finalle}$ obtenus par l'étape précédente est comparer avec un seuil déterminé expérimentalement pour détecter les transitions de diapositives entre les plans consécutives d'une vidéo par la condition suivante [31] :

$$Transition_{detecter} = (vrais\ si : LD(i-1) < \sigma\ et\ LD_i > \sigma\ et\ LD(i+1) < \sigma\ faux\ sinon) \quad (3.15)$$

Où,

- $LD(i-1)$: Est la différence de couleur CIELAB de plan précédent au plan courant.
- LD_i : Est la différence de couler CIELAB de plan courant au suivant.
- $LD(i+1)$: Est la déférence de couler CIELAB de plan suivant au dernier.

- ∂ La valeur de seuil global : basé sur des seuils utilisent le même seuil prédéfini surune vidéo entière, où la valeur est généralement définie de manière empirique.

La phase de détection permet de détecter les transitions entre les trois plans Consécutifs : précédent courant et suivant pour les deux types de vidéos de conférence : vidéo compose d'une diapositive seulement et vidéo compose une diapositive et d'une vue d'interlocuteur. Le system de détection des transitions enregistre la transition détectée entre les trois plans consécutifs par rapport à l'index du plan courant. L'organigramme suivant illustre la phase de détection :

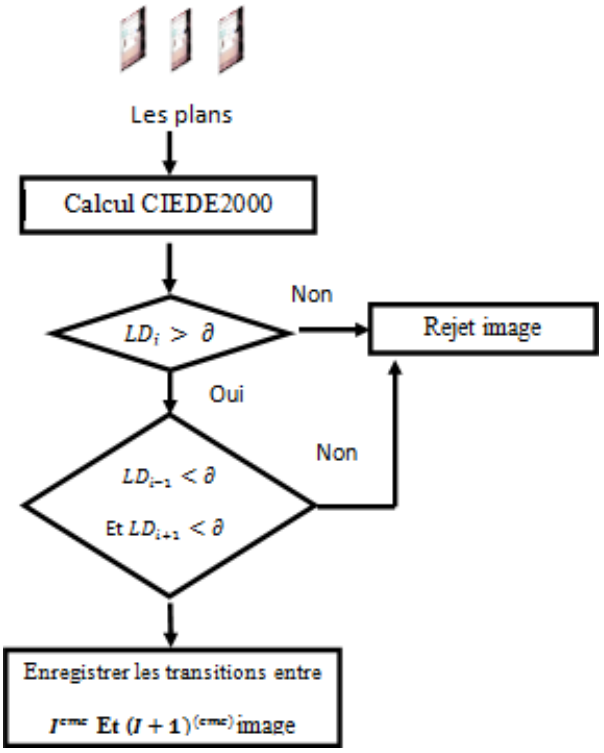


FIGURE 3.7 – Organigramme de la phase de détection.

Un Exemple de détection de transition de diapositive entre les deux plans consécutifs est présentépar la figure suivante 3.8 :

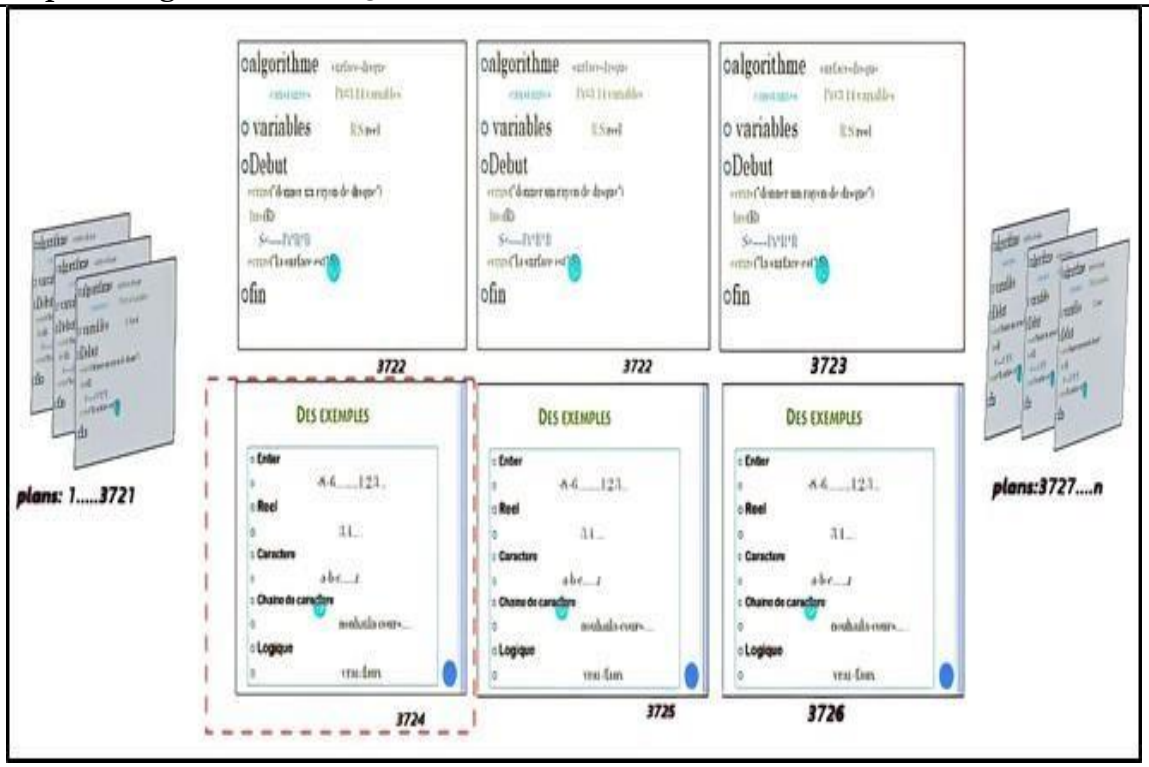


FIGURE 3.8 – Un exemple des résultats de la phase de détection : détection d'une transition de diapositive entre les deux plans consécutifs 3723,37234

3.4 Conclusion

Dans ce chapitre, nous avons présenté en détail l'architecture générale du système de détection des transitions des vidéos à usage éducatif. Pour chaque phase, nous présentions les méthodes et techniques utilisées ainsi que la démarche à suivre pour aboutir aux résultats. L'implémentation du système ainsi que les résultats obtenus seront présentées et discutées dans le chapitre suivant.

CHAPITRE 4

RESULTATS ET DISCUSSIONS

Sommaire

4.1 Introduction 35
4.2 Environnement de développement 35
4.2.1 Environnement matériel 35
4.2.2 Environnement logiciel 35
4.3 Base de données utilisée 35
4.3.1 La collection des vidéos 35
4.3.2 L'annotation de la base 37
4.4 Métriques d'évaluation 37
4.5 Résultats et discussions 38
4.5.1 Paramètres de configuration 38
4.5.2 Evaluation et discussions 40
4.5.3 Comparaison avec d'autres méthodes 41
4.6 Conclusion 42

4.1 Introduction

Dans les chapitres précédents, on a exposé en détail notre système de détection des points de transition ainsi que les différentes étapes qui le constituent. À cet effet, dans ce chapitre, on va présenter l'implémentation du système proposé. Nous commençons par décrire l'environnement de développements matériels et logiciels ainsi que la base de données utilisées. Ensuite, nous présentons et discutons les résultats obtenus ainsi qu'une comparaison avec d'autres méthodes.

4.2 Environnement de développement

4.2.1 Environnement matériel

Le traitement des vidéos est un processus gourmand en temps et en calcul. En particulier, la détection des points de transition nécessitée une comparaison en paire entre les plans adjacents précédés par un prétraitement par conséquent nécessite performances matérielles et logicielles. Le service google Collab est un outil permet d'utiliser un environnement de notebook 'Jupyter' en ligne et gratuit, qui ne nécessite aucune configuration et permet d'accéder gratuitement àdes ressources informatiques, dont des GPU.

4.2.2 Environnement logiciel

Durant l'implémentation, nous avons utilisé plusieurs bibliothèques en l'occurrence :

— **Module OpenCv** : ce module nous permet de traiter des images et des vidéos pour identifier des objets, des visages ou même l'écriture d'un humain. [1]
— **Module Numpy** : ce module fournit un objet tableau multidimensionnel hautes performances et des outils pour travailler avec ces tableaux, ces opérations peuvent être combinées avec le module Opencv [2]
— **Module Scipy** : Ce module nous permet utile pour nombreuses équations et algorithmes mathématiques. [3]

4.3 Base de données utilisée

4.3.1 La collection des vidéos

Dans ce travail, on a créé une base de données qui enveloppe une quinzaine de vidéos de conférence. Deux types de vidéos de cours sont choisis comme entrée pour l'expérimentation, des vidéos du type 1 composée uniquement de vue diapositive et celles du type 2 composées d'une vue diapositive et d'une vue présentatrice. La figure 4.1 montre l'exemple de chaque type de vidéo. Les vidéos des deux types présentent différentes dégradations afin d'évaluer notre systeme à savoir :

— Manque de luminance
— Moyens non professionnels de capture
— Mauvaise qualité de résolution durée de la vidéo
— Les informations difficiles à décerner à œil nu.

1. https ://www.geeksforgeeks.org/opencv-python-tutorial/
2. https ://www.geeksforgeeks.org/python-numpy/
3. https ://github.com/scipy/scipy

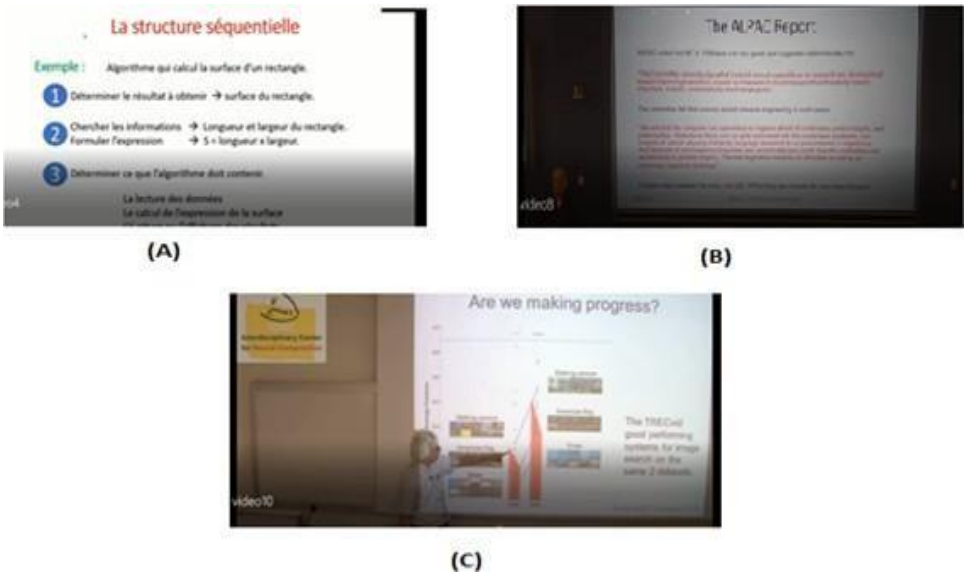


FIGURE 4.1 – Exemples des plans des vidéos de la base de données (a) vidéo de type 1 (b) et (c)vidéo de type 2.

4.3.1.1 Vidéos de type 1

Ce type contient sept vidéos caractérisées par une vue de diapositive. Les caractéristiques de chacune de ces vidéos sont données dans le tableau 4.1

Vidéos	Nombre de plans	Nombre de transitions	Durée(sec)	Plan par seconde
Video1	9738	11	973.8	10
Vidéo2	24344	22	973.973	25
Video3	18829	16	627	30
Vidéo4	18409	21	613.63	30
Video5	41155	26	1371.83	30
Video6	18339	32	1222.6	15
Video7	74681	61	2460	30

TABLE 4.1 – Caractéristiques des vidéos du type 1.

4.3.1.2 Vidéos de type 2

Ce type contient sept vidéos, trois vidéos se caractérisent par une vue de diapositive et une vue d'interlocuteur et le reste des vidéos se caractérisent aussi par une vue diapositive et un mouvement d'interlocuteur. Les caractéristiques de chacune de ces sept vidéos sont données dans le tableau 4.2

Vidéos	Nombre de plans	Nombre de transitions	Durée(sec)	Plan par seconde
Video8	25847	54	862.42	29.97
Video9	53671	44	2238.22	23.97
Video10	80815	51	3370.65	23.97
Video11	71449	51	2977	23.97
Video12	65490	51	2619.6	24
Video13	26806	33	1072.24	25
Vidéo14	14348	17	478.744	29.97

TABLE 4.2 – Caractéristiques des vidéos du type 2.

4.3.2 L'annotation de la base

L'annotation de notre base est une phase préliminaire qui consiste à faire une extraction des plans et des caractéristiques de chaque vidéo. Pour chaque vidéo, l'extraction consiste à :

1. L'analyse manuelle pour situer l'emplacement des transitions dans chaque vidéo ;
2. Enregistrer ces emplacements dans le fichier d'annotation

Les fichiers résultants seront utilisés ultérieurement pour comparaison avec les résultats détecter par le système pour comparaison avec les résultats détecter par le système pour calculer les différentes mesures d'évaluation.

4.4 Métriques d'évaluation

Les performances du système proposé ainsi que les méthodes de comparaison utilisées sont évaluées sur la base du rappel, de la précision et du F1-mesuré. Les mesures sont calculées d'après les équations suivantes :

$$Précision = (Nc)/(Nc + Nf) \tag{4.1}$$

$$Rappel = (Nc)/(Nc + Nm) \tag{4.2}$$

$$F1 - mesure = (2 * Rappel * précision)/(Rappel + précision) \tag{4.3}$$

tel que :

— **Nc** : Nombre des transitions détectées par le système
— **Nm** : Nombre des transitions non détectées par le système.
— **Nf** : Nombre de fausses transitions détectées par le système

Pour toute l'expérimentation le rappel, la précision et le F1-mesurent sont notés **Rap, Pre et F1** respectivement. Dans notre approche le facteur F1-mesure nous renseigne sur la performance du système proposé. La figure 4.2 présente les différentes valeurs du F1-mesure obtenus pour chaque vidéo.

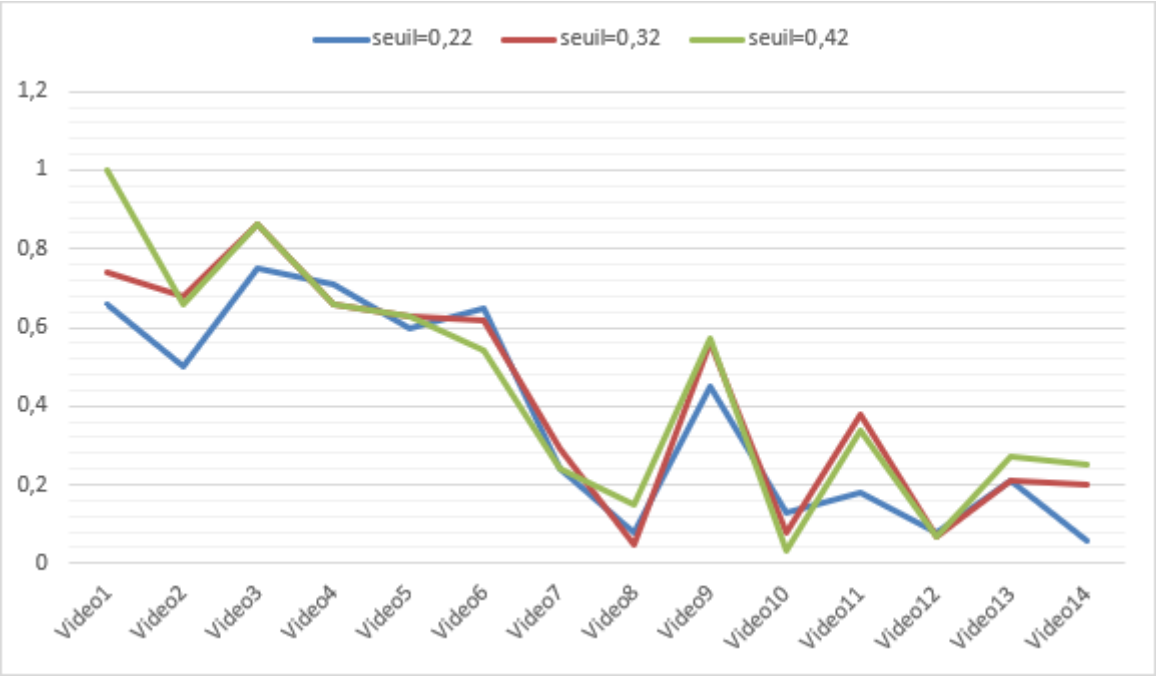


FIGURE 4.2 – Mesure F1 pour différents seuils.

4.5 Résultats et discussions

Dans cette partie, nous présentons et nous discutons les différents résultats par chaque phase du système proposé.

4.5.1 Paramètres de configuration

4.5.1.1 La phase de prétraitement

La détection des transitions nécessite une amélioration, extraction des caractéristiques et calcul de distance entre les plans adjacents ce qui impose un grand temps de traitement. Ainsi, pour minimiser le temps et les bruits on a utilisé un saut de plans. Le traitement prend en considération les plans à un intervalle de 40 plans. La valeur 40 est estimée par expérimentations pour garantir à la fois la détection des transitions et évitez les calculs inutiles. Chacun de plans est prétraité par la méthode de suppression de brume 4.3 et transformé en espace de couleur Cielab.

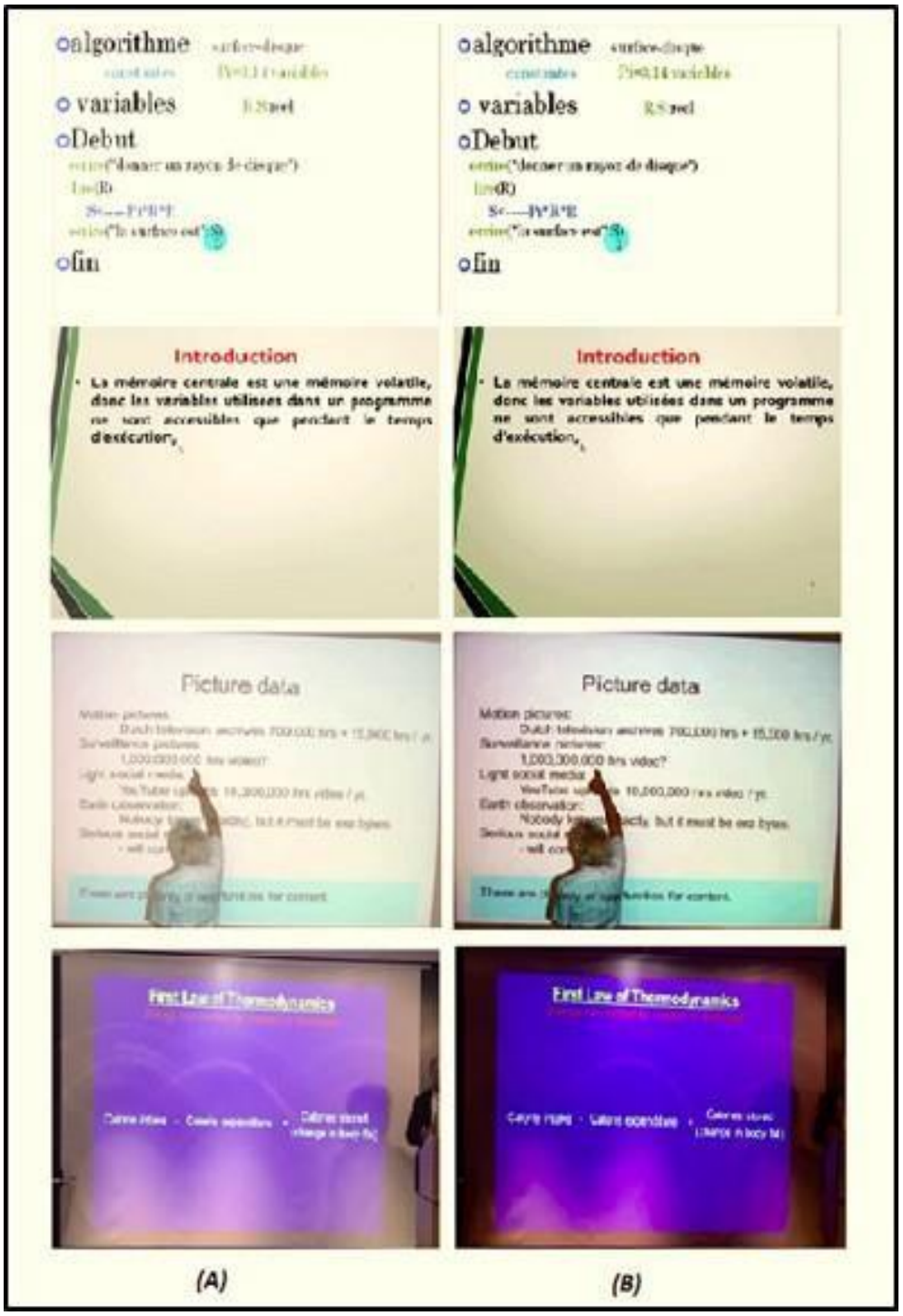


FIGURE 4.3 – Exemples de résultats de suppression du brume(A) plan original, (B)plan après la suppression de la brume.

4.5.1.2 La phase d'analyse

La division du plan en sous plans permet de mieux prendre en considération les changements locaux de la couleur. Le nombre de devisions possibles peut varier entre 2 à 10. Pour cela, nous avons expérimenté notre approche en utilisant 4,6 et 8 sous plans les figures (4.4) et (4.5) présentent les résultats obtenus

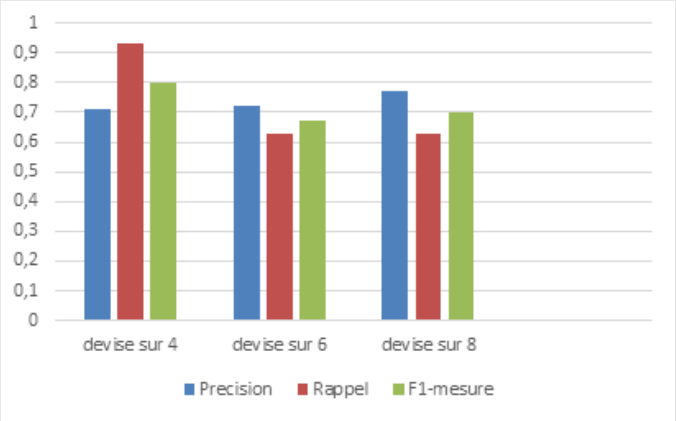


FIGURE4.4 – Comparaison des performances de détection sur la vidéo 1 pour différents nombres de sous-plans

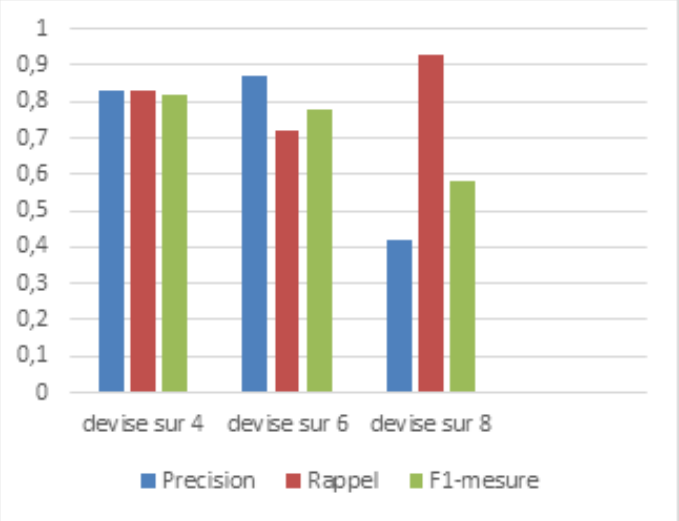


FIGURE 4.5 – Comparaison des performances de détection sur la vidéo 9 pour différents nombres de sous-plans

La phase de détection

La phase de détection utilise un seuil global, les performances du système sont fortement liées à cette valeur qui doit permettre de détecter toutes les transitions et moins de fausses détections. Pour les expérimentations trois valeurs ont étés utilisées à savoir 0.22, 0.32 et 0.42 choisies à partir des histogrammes de différence entre les plans. Chaque histogramme contient des pics minimaux qui représentent la présence d'une transition, ces pics commencent à partir de 0.20. Deux histogrammes sont donnes dans les figures 4.5 et 4.6.

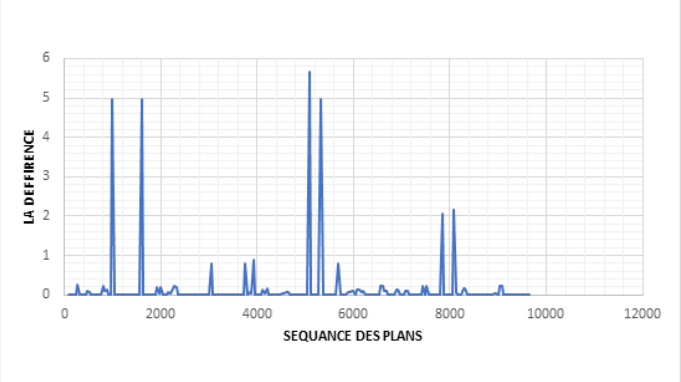


FIGURE 4.6 – La différence CIEDE2000 entre plans de la vidéo 1

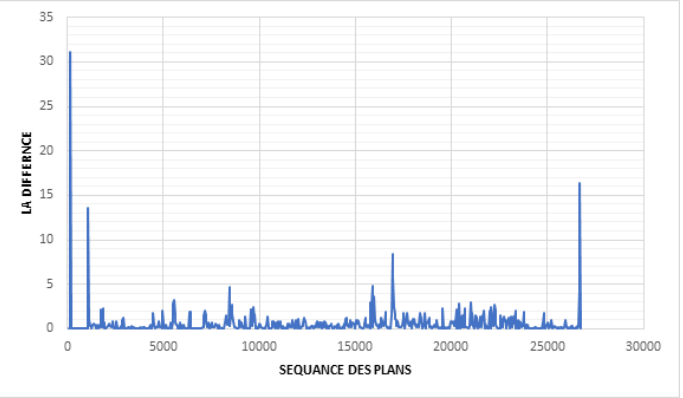


FIGURE 4.7 – La différence CIEDE2000 entre plans de la vidéo 11

4.5.2 Evaluation et discussions

4.5.2.1 Phase de détection

Pour l'évaluation de la phase de détection, nous avons obtenu les différents résultats suivants :

Les résultats présentés nous permettent de conclure que la précision pour le type 1 que celle des vidéos du type 2 qui sont caractérisées par la présence de l'interlocuteur ce qui influe négativement sur la focalisation du système sur le contenu de la diapositive. On remarque aussi pour le même type il y a une variation entre les résultats à cause de leurs différentes qualités, l'emplacement de la caméra et l'influence de la lumière. La figure suivante montre les plans avant et après la suppression de la brume. Les deux figures suivantes montrent deux diagrammes de résultats des vidéos pour chaque type.

Type des vidéos	Rappel	Précision	F-mesure
Type 1	0.58	0.7	0.63
Type 2	0,23	0,19	0,21

TABLE 4.3 – les moyennes de précision, de rappel et de F1-mesure pour chaque type de vidé

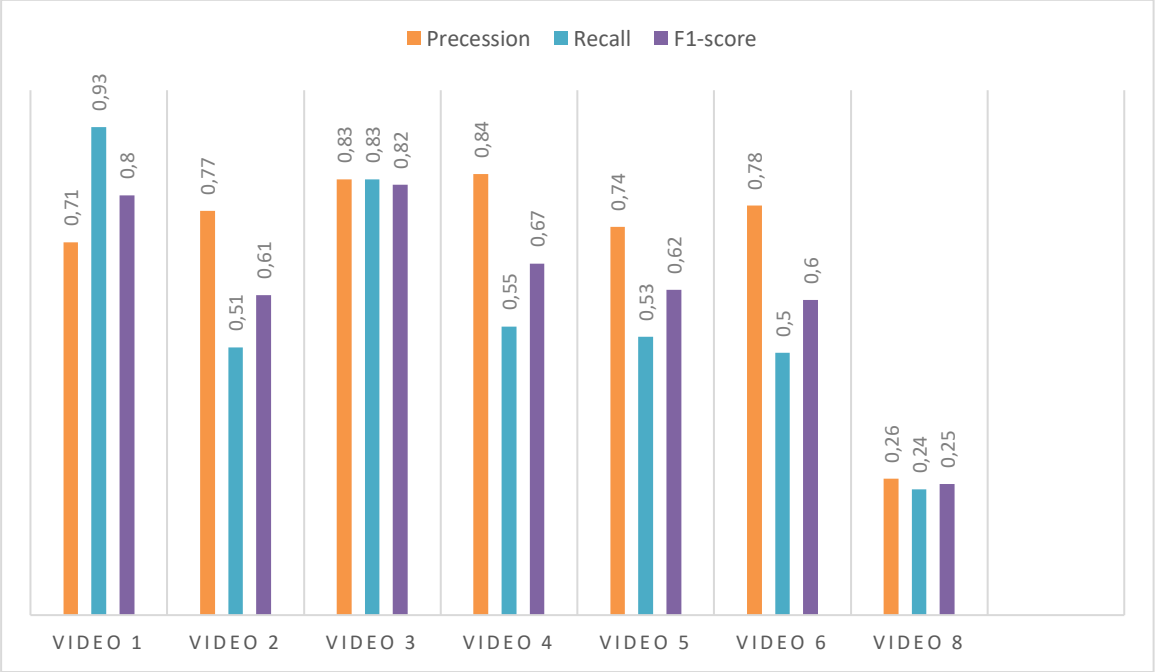


FIguRE 4.8 – Evaluation des performance sles vidéos de type 1

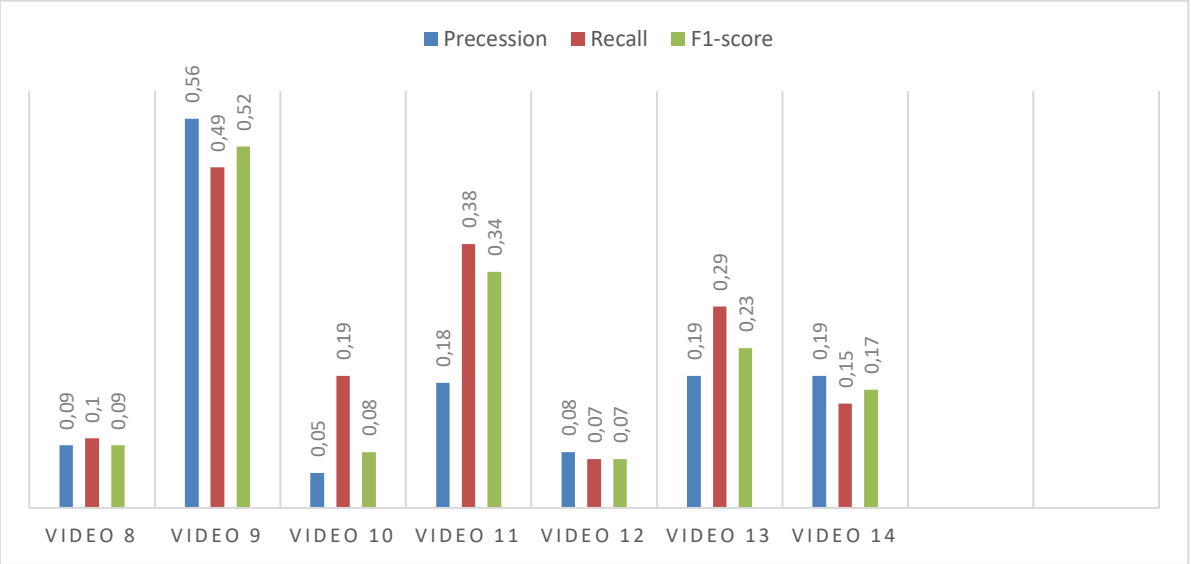


FIGURE 4.9 – Evaluation des performancessur les vidéos de type 2

4.5.3 Comparaison avec d'autres méthodes

Pour avoir une évaluation globale sur la fiabilité de notre système proposé, nous avons comparé nos résultats avec ceux d'autres méthodes pour les différentes phases.

4.5.3.1 Comparaison avec autres méthodes d'amélioration

Le choix d'une méthode efficace de prétraitement est de grande importance pour la qualité des résultats du système de détections de transition. Pour cela, nous avant comparait notre méthode à une méthode de double éclairage qui est base sur les algorithmes de rétines connue pour son efficacité sur les vidéos. Mais les expérimentations ont prouvé que les vidéos de conférences subissent des illuminations qui provoquent un chèque pour la méthode comme le montre le tableau4.4.

Méthode	Rappel	Précision	F-mesure
Méthode suppression de brume	0.93	0.71	0.8
Méthode double éclairage	0.8	0.24	0.74

TABLE 4.4 – Example sur les valeurs de performances de système entre la méthode de suppression de la brume et la méthode de double éclairage pour le vidéo 1 .

4.5.3.2 Comparaison avec autres méthodes de détection

On a comparé le système proposé par une méthode de base 'histogramme de couleur' la plus utilisée dans les systèmes de segmentation et par la méthode de pixel global qui considérer meilleure pour les résultats qu'elle donne [13]. Le tableau suivant montre un état comparatif de la méthode proposée par rapport à celles du pixel global et histogramme de couleur. Les deux tableaux suivants montrent la moyenne des résultats pour les trois méthodes de comparaison :

Méthode	Rappel	Précision	F-mesure
Systeme proposé	0.5	0.7	0.63
PG	0.7	0.8	0.74
Hist	0.09	0.13	0.1

TABLE 4.5 – les moyennes de précision, de rappel et de F1-mesure des vidéos de Type-1 pour chaque méthode.

Méthode	Rappel	Précision	F-mesure
Systeme proposé	0,23	0,19	0,21
PG	0.56	0.18	0.23
Hist	0.19	0.17	0.10

TABLE 4.6 – les moyennes de précision, de rappel et de F1-mesure des vidéos de Type-2 pour chaque méthode.

4.5.3.3 Comparaison en différents espaces de couleur

Notre systeme utilise l'espace CIElab pour les différents traitements, Nous avons également comparé notre systeme en espace de couleurs Rouge vert-bleu. La comparaison montre que le système proposé donne des meilleurs résultats par rapport à l'espace colorimétrique RVB

Espace	Rappel	Précision	F-mesure
CIElab	0.5	0.7	0.63
RVB	0.09	0.13	0.1

TABLE 4.7 – Etat comparatif de la méthode proposée en espaces de couleur RVB et LAB pour le type 1

Les espaces de couleurs	Rappel	Précision	F-mesure
LAB	0,23	0,19	0,21
RVB	0.19	0.17	0.10

TABLE 4.8 – Etat comparatif de la méthode proposée en espaces de couleur RVB et LAB pour le type 2

4.6 Conclusion

Dans ce chapitre, nous avons présenté l'implémentation et les résultats pour chaque phase du système proposé. Les résultats pour chaque vidéo ont était présentés, comparer et discuter.

CHAPITRE 5

CONCLUSIONS GENRALE

Sommaire

5.1 Conclusion 44
5.2 Perspectives 44

5.1 Conclusion

Dans ce mémoire, nous avons présenté les systèmes de segmentation temporelle des vidéos de conférences. Nous avons décrit les différentes phases constituant ce système : extraction des caractéristiques visuelles, mesure de similarité et la détection. Pour une meilleure compréhension du problème et des difficultés des vidéos de conférence nous avons basé sur les caractéristiques visuelles des plans et leurs points de transition. Après avoir cerné les problèmes, on était en mesure de proposer une solution efficace après avoir présentés les méthodes et techniques les plus utilisés et qui permettent le prétraitement, l'analyse et la détection des points de transition. Nous avons proposé un système de détections de transition composé de plusieurs phases : le prétraitement qui consiste l'amélioration des plans pour la phase d'extraction des caractéristiques visuelles, et pour les mesures de similarité nous avons utilisé la formule de déférence d'couleurs CIEDE2000 à base d'un seuil global. Les résultats qu'on a obtenus nous permettions de conclure que : la suppression de la brume est une méthode efficace pour l'amélioration des plans et la formule de déférence d'couleurs CIEDE2000 permet une grande précision sur l'espaceLAB. Pour la détection les valeurs des performances du type 1 des vidéos sont plus élevées que les résultats de type 2.

5.2 Perspectives

Les méthodes proposées dans ce mémoire ont montré des bonnes performances. En perspective, on prévoit l'amélioration du système de segmentation temporelle des vidéos de conférence en fonction d'élimination de bruit avec moins de complexité de calcul. On envisage aussi de travailler sur l'estimation du seuil.

BIBLIOGRAPHIE

[1] ALLOGNON Odilon, TOURÉ Fodé, ET AÏMEUR Esma. “Learning by challenging: A social network and privacy based approach”. In: International Conference on Education and e-Learning Innovations. IEEE, 2012. p. 1-6.

[2] MASON, Robin et RENNIE, Frank. Livre: ‘’Elearning: The key concepts”. Routledge, 2006.

[3] IKKOU, Lahoucine, BERDI, Abdelaziz, et SEGHYAR, Nabil.” La formation e-Learning dans l’entreprise: étude empirique auprès des banques marocaines. Revue Internationale d’Economie Numérique” in : 2020, vol. 2, no 2, p. 11-21.

[4] BALANCIER, Pascal, GEORGES, François, JACOBS, Sarah, et al. ”L'e-Learning dans l'enseignement supérieur”. Environnement International Francophone. 2006.

[5] KUMAR BASAK, Sujit WOTTO Marguerite, et BELANGER Paul.” E-learning, M-learning and D-learning: Conceptual definition and comparative analysis”. In: Elearning and Digital Media, 2018, vol. 15, no 4, p. 191-216.

[6] NOOR-UL-AMIN, Syed. ‘’An effective use of ICT for education and learning by drawing on worldwide knowledge, research, and experience”. In: ICT as a Change Agent for Education. India: Department of Education, University of Kashmir, 2013, p. 1-13.

[7] RAJ, S. Anthony Philomen, et al. ‘’A Comprehensive Study on Learning Styles, E Learning Model and Tools’’. in :Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, vol. 12, no 10, p. 2504-2510.

[8] CHANDRAN, Daniel et KEMPEGOWDA, Sunil. ‘’Hybrid E-learning platform based on cloud architecture model: A proposa”l. In: 2010 International Conference on Signal and Image Processing. IEEE, 2010. p. 534-537.

[9] FLANK, Sharon. ‘’ Multimedia technology in context”.in: IEEE MultiMedia, 2002, vol. 9, no 3, p. 12-17.

[10] GHIRARDINI Beatrice, LANDRISCINA F. et SHAPIRO B livre : Méthodologies pour le développement de cours E-learning. Un guide pour concevoir et élaborer des cours d’apprentissage numérique. In: FAO, 2011.

[11] BAGDANOV, Andrew D., BERTINI, Marco, DEL BIMBO, Alberto, et al. ‘’Semantic annotation and retrieval of video events using multimedia ontologies”. In: International Conference on Semantic Computing (ICSC 2007). IEEE, 2007. p. 713-720.

[12] HAMPEL, Regine et HAUCK, Mirjam. ‘’Towards an effective use of audio conferencing in distance language courses”. In: Language Learning & Technology, 2004, vol. 8, no 1, p. 66-82.

[13] ERUVARAM, Purushotham, RAMANI, Kasarapu, ET BINDU C. Shoba. ” An experimental comparative study on slide change detection in lecture videos”. In: International Journal of Information Technology, 2020, vol. 12, no 2, p. 429-436.

[14] RADHA, N. ‘’Video retrieval using speech and text in video”. In: 2016 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2016. p. 1-6.

[15] DALLAS, Patrick S. ‘’Video Conferencing application to distance education with particular reference to small states” in: Proceedings of the Distance Education in Small States, 2001, p. 291-298.

[16] AMIRI, Ali et FATHY, Mahmood. ‘’Video shot boundary detection using QR-decomposition and gaussian transition detection”.in: EURASIP Journal on Advances in Signal Processing, 2010, vol. 2009, p. 1-12.

[17] ABDULHUSSAIN, Sadiq H., RAMLI, Abd Rahman, SARIPAN, M. Iqbal. ‘’Methods and challenges in shot boundary detection: A review”. In: Entropy, 2018, vol. 20, no 4, p. 214.

[18] YEUNG, Minerva M. et YEO, Boon-Lock. ‘’Time-constrained clustering for segmentation of video into story units”. In : Proceedings of 13th International Conference on Pattern Recognition. IEEE, 1996. p. 375-380.

[19] BLANK, Moshe, GORELICK, Lena, SHECHTMAN, Eli, et al. ‘’Actions as space-time shapes”. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005. p. 1395-1402.

[20] WATKINSON, John.livre: ‘’The art of digital video”. In: Routledge, 2013.

[21] SAFADI, Bahjat. *Indexation sémantique des images et des vidéos par apprentissage actif*. 2012. Thèse de doctorat. Université de Grenoble.

[22] CHAN, Shermann S.-M., LI, Qing, WU, Yi, et al. ‘’A Hybrid Approach to Video Retrieval in a Generic Video Management and Application Processing Framework”. In: ICME. 2001.

[23] SCUTURICI, Mihaela. ‘’Contribution aux techniques orientées objet de gestion des séquences vidéo pour les serveurs Web”.in : 2002. Thèse de doctorat. Thèse de Doctorat en Informatique. Institut National des Sciences Appliquées de Lyon.

[24] HAIDAR, Siba. ‘’Comparaison des documents audiovisuels par matrice de similarité”.in : 2005. Thèse de doctorat. Université Paul Sabatier-Toulouse III.

[25] BOURENANE, Mohamed Amine. ‘’Un outil pour l'indexation des vidéos personnelles par le contenuin” .in : 2009. Thèse de doctorat. Université du Québec à Trois-Rivières.

[26] RIAZ, Muhammad, KANG, Gwangwon, KIM, Youngbae, et al. ‘’Efficient image retrieval using adaptive segmentation of HSV color space” In: 2008 International Conference on Computational Sciences and Its Applications. IEEE, 2008. p. 491-496.

[27] RAHMAT, Romi Fadillah, CHAIRUNNISA, Tengku, GUNAWAN, Dani, et al. ‘’Skin color segmentation using multi-color space threshold”. In: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS). IEEE, 2016. p. 391-396.

[28] LY, Bao Chau K., DYER, Ethan B., FEIG, Jessica L., et al. ‘’Research techniques made simple: cutaneous colorimetry: a reliable technique for objective skin color measurement”. Journal of Investigative Dermatology, 2020, vol. 140, no 1, p. 3-12. e1.

[29] LORESCO, Pocholo James M., VALENZUELA, Ira C., et DADIOS, Elmer P. ‘’Color space analysis using KNN for lettuce crop stages identification in smart farm setup”. In : TENCON 2018-2018 IEEE Region 10 Conference. IEEE, 2018. p. 2040-2044.

[30] JEONG, Hyun Ji, KIM, Tak-Eun, KIM, Hyeon Gyu, et al. ‘’Automatic detection of slide transitions in lecture videos. Multimedia Tools and Applications”, 2015, vol. 74, no 18, p. 7537- 7554.

[31] CHAKRABORTY, Saptarshi, THOUNAOJAM, Dalton Meitei, et SINHA, Nidul. ‘’A Shot boundary Detection Technique based on Visual Colour Information”. In Multimedia Tools and Applications springer, 2021, vol. 80, no 3, p. 4007-4022.

[32] SHARMA, Gaurav, WU, Wencheng, ET DALAL, Edul N. ‘’The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations”. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 2005, vol. 30, no 1, p. 21-30

[33] LUO, M. Ronnier, CUI, Guihua, et RIGG, Bryan. ‘’The development of the CIE 2000 colour-difference formula: CIEDE2000”. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 2001, vol. 26, no 5, p. 340-350..

[34] SHARMA, Gaurav, WU, Wencheng, DALAL, Edul N., et al. ‘’Mathematical discontinuities in CIEDE2000 color difference computations”. In: Color and Imaging Conference. Society for Imaging Science and Technology, 2004. p. 334-339.

[35] NGO, Dat, LEE, Seungmin, NGUYEN, Quoc-Hieu, et al. ‘’Single image haze removal from image enhancement perspective for real-time vision-based systems”. In Sensors, 2020, vol. 20, no 18, p. 5170.

[36] PARK, Yongmin et KIM, Tae-Hwan. ‘’Fast execution schemes for dark-channel-prior-based

outdoor

video dehazing”.in: IEEE Access, 2018, vol. 6, p. 10003-10014.

[37] ZHANG, Qing, NIE, Yongwei, et ZHENG, Wei-Shi. Dual illumination estimation for robust exposure correction. In : *Computer Graphics Forum*. 2019. p. 243-252.

[38] AL-HASHIM, Mohammad Abid et AL-AMEEN, Zohair. Retinex-Based Multiphase Algorithm for Low-Light Image Enhancement. *Traitement du Signal*, 2020, vol. 37, no 5.