

Découpage automatique de la vidéo en plans

A.MAREDJ, F.SAADI, D.MEDDOUR

Laboratoire Bases de Données et Systèmes d'information

CE.R.I.S.T

Rue des trois frères Aïssou, Ben Aknoun Alger, Algérie.

E-mail : amaredj@mail.cerist.dz

1. Introduction

Aujourd'hui, l'information audiovisuelle est utilisée dans des secteurs de plus en plus nombreux de la vie courante et joue un rôle très important dans notre société moderne. Ceci explique le besoin de structurer cette information pour faciliter sa recherche et l'accès à son contenu. Le problème de la consultation des documents audiovisuels réside dans la visualisation séquentielle de leur contenu pour pouvoir localiser un passage particulier auquel on s'intéresse.

L'amélioration des services offerts aux utilisateurs des bases de données vidéos doit passer nécessairement par l'introduction de l'accès direct au contenu de ces documents. Il faut alors découper la vidéo en parties (plans) significatives auxquelles l'utilisateur aura accès.

Les plans représentent l'équivalent d'une table des matières permettant un accès rapide et guidé au contenu du document. Elle constitue une sorte de résumé visuel du document vidéo.

Le découpage de la vidéo en plans (ou segmentation temporelle) peut être mené par un opérateur humain, mais seule son automatisation peut constituer une alternative sûre et rapide et ce en palliant aux problèmes inhérents à l'approche manuelle (lenteur, voire impossibilité de traitement lorsqu'il s'agit de Base de Données volumineuses).

Une des difficultés rencontrée avec les techniques utilisées jusqu'à présent pour la détection des plans réside dans leur utilisation d'un seul seuil ce qui les rend très limitatives pour la détection des différents types de transitions. En effet, l'évolution temporelle de la différence des histogrammes des images successives présente des pics très prononcés lors des cuts, et des valeurs modérées lors des transitions progressives présentes sur l'étendue temporelle de la transition. Ce qui rend la détection des changements de plans et la distinction entre les cuts et les transitions progressives très aléatoire en fonction de la valeur du seuil fixée.

Dans cet article nous présentons une technique de détection de plans vidéo à deux seuils. L'objectif de la technique proposée est de pouvoir détecter avec efficacité les différents types de transitions (cuts, fondu et volet). Nous commençons par la définition d'un plan ainsi que les différentes techniques utilisées en montage vidéo pour recorder les plans. Un rappel sur les différentes techniques de détection de plans sera suivi d'une discussion. Par la suite, nous présentons en détail la technique à deux seuils pour la détection de changement de plans ainsi que les résultats expérimentaux. Nous terminerons par une conclusion.

2. Plans

Le découpage en plans (ou segmentation temporelle) est la première étape à effectuer pour l'analyse de la vidéo. Il constitue l'étape la plus importante en conditionnant toute réussite des traitements ultérieurs sur la vidéo.

Un plan est une collection d'images successives dont le contenu est homogène (fig1). Le plan est le composant le plus élémentaire dans l'opération de montage d'une vidéo.

Il s'agit alors de retrouver la structure temporelle résultant de l'opération de montage du document. Ce montage consiste, pour simplifier, à concaténer des

extraits (plans) de bande vidéo ou de film que le réalisateur aura retenu.

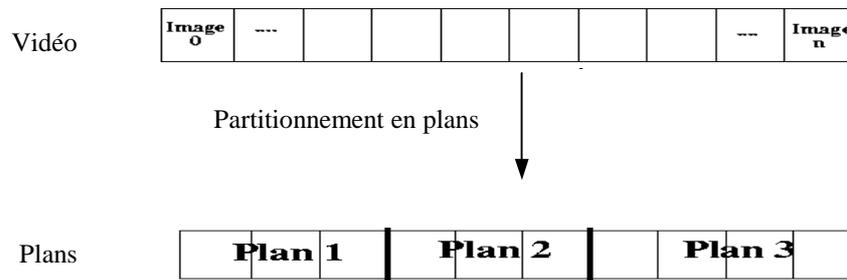


Fig-1 *Structure interne de la vidéo*

3. Montage vidéo et raccord vidéo

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

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

3.1. Cuts

Le raccord le plus simple et le plus fréquent entre deux plans est le "cut" par lequel on passe sans transition d'un plan au suivant (fig 2. a). Dans un film il traduit un changement de situation radicale où il n'y a aucune relation entre deux plans.

3.2. Transitions progressives

Ces raccords consistent le plus souvent à passer progressivement d'une image du premier plan à une image du second plan et sont alors localisés sur une étendue variable d'images.

Si les transitions possibles sont très variées, la plupart des effets rencontrés peuvent être classés en deux catégories : fondus et volets ou en être une combinaison.

3.2.1. fondu

Le *fondu* consiste à passer progressivement d'une séquence d'images $I_1(t)$ à une autre séquence $I_2(t)$ en combinant les images de ces deux séquences pendant un intervalle de temps prédéfini (fig 2. b). En pratique $I_1(t)$ et $I_2(t)$ peuvent évoluer pendant la transition (en raison d'un mouvement de caméra, par exemple). Cette combinaison alloue une contribution variable $\tau(t)$ de chaque séquence de la transition progressive en terme d'intensité selon une loi généralement linéaire par rapport au temps. L'intensité de l'image composée s'exprime comme suit [BOU96] :

$$I(t) = \tau(t) I_1(t) + (1 - \tau(t)) I_2(t) \quad (1)$$

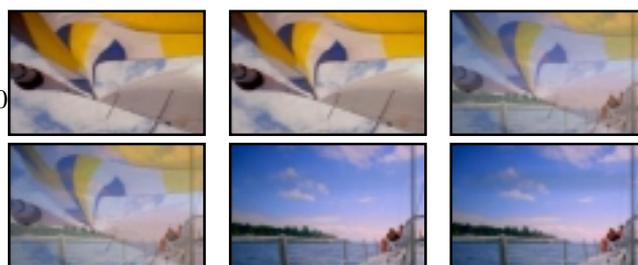
3.2.2. Volet

Le *volet* consiste à passer progressivement d'un plan $I_1(t)$ à un autre plan $I_2(t)$ en variant dans le temps la proportion en terme d'espace dans l'image allouée à chaque plan (fig 2. c). Il peut être effectué selon un déplacement horizontal, vertical ou plus complexe (rotation...). Cette transition peut par exemple produire un effet de page que l'on tourne, ce qui peut expliquer son utilisation abondante dans les journaux télévisés.



a)

RIST VOL.11 N° 0



4. Les Techniques utilisées

Nous trouvons dans la littérature deux grandes approches pour la structuration de la vidéo en plans : celles qui procèdent par un découpage de la vidéo, auquel cas on s'attache à détecter les changements de plans, et celles qui procèdent par un regroupement (clustering) des images appartenant à un même plan.

La première exploite la différence entre les images successives pour détecter le changement de plans. La seconde exploite la similarité des images (successives) pour les regrouper en Clusters (groupes) d'images cohérentes appartenant à un même plan.

L'avantage principal de la première technique par rapport à la seconde réside dans la rapidité de traitement car pour cette approche on doit analyser, la plus part du temps, une différence entre deux images successives, alors que pour la seconde approche on doit garder trace du groupe d'images auquel appartient l'image pour analyser la similarité entre l'image en cours et les images du plan.

La technique la plus simple de détection de changement de plans reste la comparaison pixel à pixel des images successives [KAS98][KEN98]. Une différence d'intensité maximum tolérée entre pixels correspondants (qui ont les mêmes coordonnées) des images successives est choisie, puis le nombre de pixels considérés comme différents entre instants successifs est calculé. Un seuillage sur cette quantité permet de décider de la présence ou non d'une transition. Les performances de cette technique sont assez rapidement dégradées par des mouvements de caméra ou des scènes de forte activité. Une telle méthode peut être appropriée dans certains contextes applicatifs (tournage avec caméra fixe par exemple, contenu d'image "maîtrisé" au sens où on connaît les phénomènes perturbant le découpage en plans).

Une autre technique repose sur l'utilisation d'histogrammes de couleur ou d'intensité [BOU97]. Ces histogrammes sont généralement évalués de manière globale sur l'image, et varient peu lors de mouvements modérés d'objets ou de la caméra. Dans ce cas on

utilise alors un simple seuillage sur la différence des histogrammes d'images successives, ou un test sur la vraisemblance (essentiellement le test du KHI-DEUX χ^2) pour affirmer que les deux histogrammes proviennent de la même distribution ou non.

Diverses techniques intermédiaires entre la comparaison pixel à pixel et la comparaison d'histogrammes globaux ont été proposées [KEN98]. L'image est alors partitionnée en blocs réguliers, un test de similarité est effectué pour chaque paire de blocs, puis un changement de plans est identifié par la présence d'un nombre suffisant de tests négatifs.

Une technique analogue consiste à utiliser à la place des histogrammes globaux des histogrammes calculés sur les lignes et les colonnes, ils préservent le caractère local et donnent ainsi plus de sens à la comparaison.

Les techniques mentionnées jusqu'ici exploitent directement l'intensité de l'image. La mesure de changement de contenu entre images successives peut se baser sur les contours extraits de l'image [CUE98]. Il s'agit alors d'évaluer la proportion des pixels de contours dans l'image $I(t)$ également présents dans l'image $I(t+1)$ et vérifiant un critère de proximité spatiale. Cette proportion est idéalement proche de 1 si $I(t)$ et $I(t+1)$ appartiennent au même plan, proche de 0 lors d'un cut, et produit un pic de moindre amplitude mais de plus longue durée lors d'une transition progressive. Une autre technique assez proche utilise la comparaison des régions extraites des images. Cette technique tente de trouver une correspondance entre les objets (régions) de deux images successives, et de déterminer le nombre d'objets disparus et apparus entre les deux instants, si ce nombre dépasse une certaine limite (calculé par rapport au nombre d'objets contenus dans les images), on considère que le contenu à changer et qu'on est en présence d'une transition.

Une technique reposant sur l'analyse du mouvement de la caméra qui correspond au mouvement dominant dans les images (estimé sur la région constituée par l'image entière), a été proposé dans [BOU97]. A l'intérieur d'un plan le mouvement de la caméra est supposé être le même. Cette technique consiste alors à estimer le

mouvement dominant sur l'étendu de la vidéo, un test permet de séparer les mouvements différents et ainsi de détecter les changements de plans.

Dans un contexte applicatif où la source vidéo est transmise par un réseau ou stockée sur support (CD-ROM), et où les exigences de rapidité de traitement sont fortes, il peut être très important de pouvoir effectuer le découpage directement à partir de la forme compressée selon les normes MPEG-1 ou MPEG-2 [CUE98]. Ce schéma de compression utilise un algorithme de détection de changement entre images orientées bloc [TEN95]. Cette technique profite de la puissance de l'algorithme de codage MPEG qui est bien connu, et offre de bons résultats. Mais, son application est liée aux fichiers compressés avec un schéma MPEG, ce qui constitue une limitation importante, et restreint son domaine d'application.

Discussion

Nous pouvons classer les techniques présentées en deux catégories en fonction de la qualité des découpages obtenus. La première regroupe les techniques, simples dans leur implémentation, qui mènent à des performances pouvant satisfaire des applications qui réclament une rapidité de traitement au détriment de la qualité des résultats (difficulté de distinguer entre les cuts et les transitions progressives) .

Dans la seconde catégorie, on retrouve celles qui donnent des résultats précis et acceptables mais qui nécessitent des temps de calcul considérables. Elles sont utilisées en amont des applications qui utilisent le découpages à des fins d'analyse de contenu (indexation par le contenu, recherche par le contenu, etc..).

Dans ce qui suit, nous allons présenter une technique où nous allons allier rapidité de traitement et précision de résultats. Pour ce faire, on utilise le principe d'histogramme globaux d'intensité pour la quantification, la comparaison est basée sur le test statistique de Kolmogorov-Smirnov et la détection des différentes transitions se fera par l'introduction d'une technique à deux seuils.

6. Technique proposée

La technique que nous proposons se base sur la comparaison d'histogrammes globaux d'intensité calculés sur les images successives. Un changement de plans sera détecté si la différence dépasse un certain seuil.

Soit $H(t)$ et $H(t+1)$ les histogrammes des images, quantifiés sur n boîtes aux instants t et $t+1$. La différence entre les deux histogrammes est exprimée par :

$$D_H = \sum_{i=1}^n |H_i(t) - H_i(t+1)| \quad (2)$$

Le changement sera détecté si D_H dépasse un seuil D_{\max} . Le problème à ce niveau reste la fixation du seuil D_{\max} , qui est le plus souvent arbitraire ou laissée aux expérimentations de l'utilisateur.

Une solution au problème réside dans l'utilisation de tests statistiques pour tester la similarité entre les deux histogrammes. Rappelons que les deux histogrammes sont considérés comme deux fonctions de répartition empiriques [TAS89]. Le test qui a été utilisé est le test du χ^2 (KHI-Deux) qui consiste à calculer la distance (dite du χ^2) :

$$D_{\chi^2} = \sum_{i=1}^n \left| \frac{(H_i(t) - H_i(t+1))^2}{H_i(t)} \right| \quad (3)$$

Le seuil du test C_{χ^2} a été tabulé et dépend de la taille de l'échantillon n (taille des images en pixels) et de la précision voulue du test (α). Cette technique a l'avantage de ne pas dépendre d'un seuil fixe, et en plus elle permet de détecter si les deux images (qui sont considérées comme la réalisation de variables aléatoires) ne sont pas dérivées de la même distribution statistique.

Mais à ce niveau nous avons relevé deux problèmes :

Le test du χ^2 n'est pas assez puissant pour détecter toutes les transitions (cuts, volets, fondus), surtout en ce qui concerne les transitions longues dans lesquelles la variation de l'intensité entre images

successives est très petite et ne peut pas être bien décelée par la distance du Khi-Deux. De plus, ce test n'est pas adapté pour le cas de distribution empirique, et dans le cas de grand échantillon.

L'utilisation d'un seul seuil est très limitative on ce qui concerne la détection des différents types de transitions. En effet, l'évolution temporelle de $\sum_i |H_i(t) - H_i(t+1)|$ présente des pics très prononcés lors des cuts, et des valeurs modérées lors des transitions progressives présentes sur l'étendue temporelle de la transition (voir figure 3). Ceci a pour effet de détecter un changement sur chaque image appartenant à la transition progressive (qui s'étale sur quelques images).

Pour pallier à ces inconvénients nous proposons ce qui suit :

Au lieu du test du Khi-Deux nous allons utiliser le test de Kolmogorov-Smirnov (K-S). Ce test est très puissant, appliqué sur des distributions empiriques il permet de les distinguer grâce à la précision de sa distance.

La détection des différents types de transitions se fait grâce à une stratégie à deux seuils qui permet de les identifier et de les distingués.

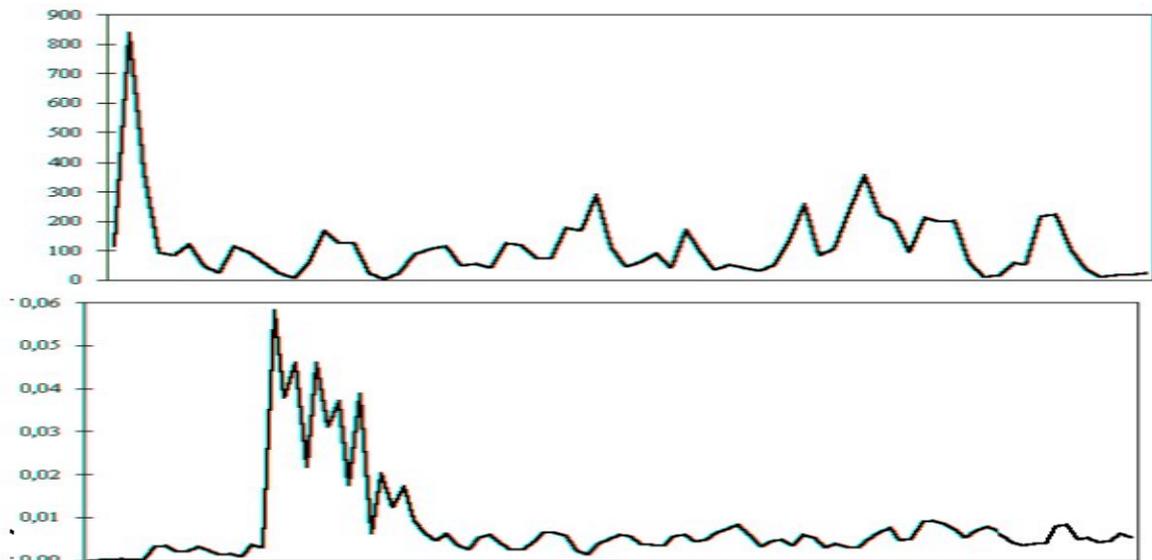


Fig-3 variation de la distance du Khi-Deux (en haut) et de la distance de K-S (en bas) entre les images successives dans un intervalle de 100 images. Comme on peut le voir la distance du χ^2 ne permet pas de détecter une telle transition.

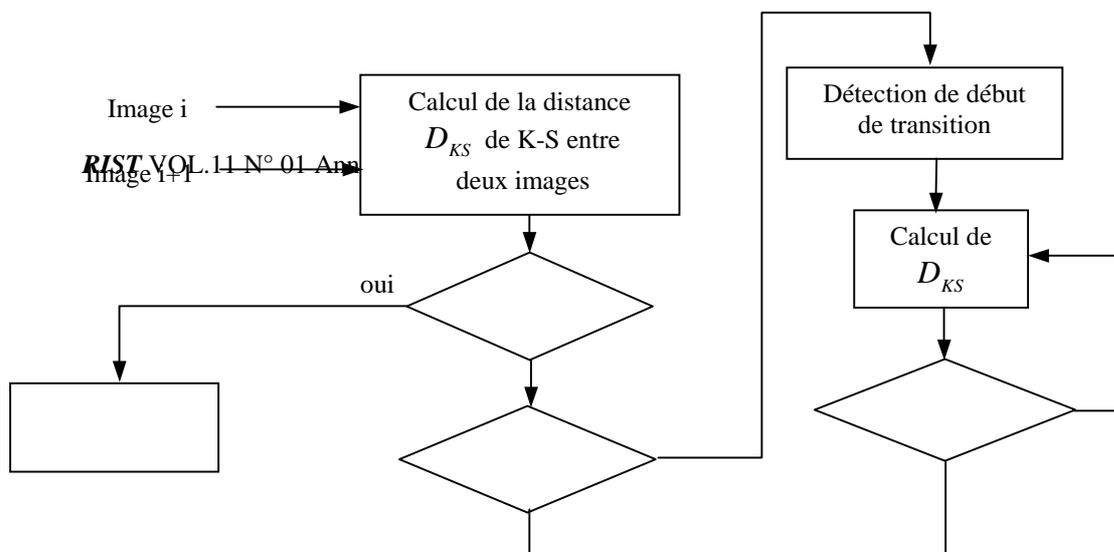
6.1 Procédure de détection de changement de plans à deux seuils

Nous allons dans ce qui suit, détailler la technique proposée pour la détection de changement de plans à deux seuils. Le principe est basé sur la séparation de la détection des cuts et des transitions progressives (fondu ou volet).

Le premier seuil S_1 permet de détecter les cuts, à cet effet, il est fixé à une grande valeur par rapport à la distance entre images successives.

Le deuxième seuil S_2 , sert à détecter la présence d'une transition progressive. La valeur de ce seuil est celle du test de komogorov-Smirnov.

L'utilisation d'une telle technique se justifie par l'analyse du graphe mesurant la distance de K-S entre images successives (voir fig 5). En étudiant l'évolution temporelle de la distance de K-S on remarque qu'elle présente des pics d'une valeur supérieure à 0.1 lors des cuts, et des variations modérées lors des transitions progressives, présentes sur l'étendue temporelle de la transition (fig 5a),. L'opération de détection de transitions progressives est déclenchée dès que la distance de K-S entre deux images successives dépasse le seuil du test de K-S (le seuil S_2). Les images suivantes seront considérées appartenant à la transitions jusqu'à ce que la distance de K-S prenne des valeurs inférieures au seuil S_2 (fig 5b) . La technique de détection de changement que nous proposons est schématisée par la figure 4.



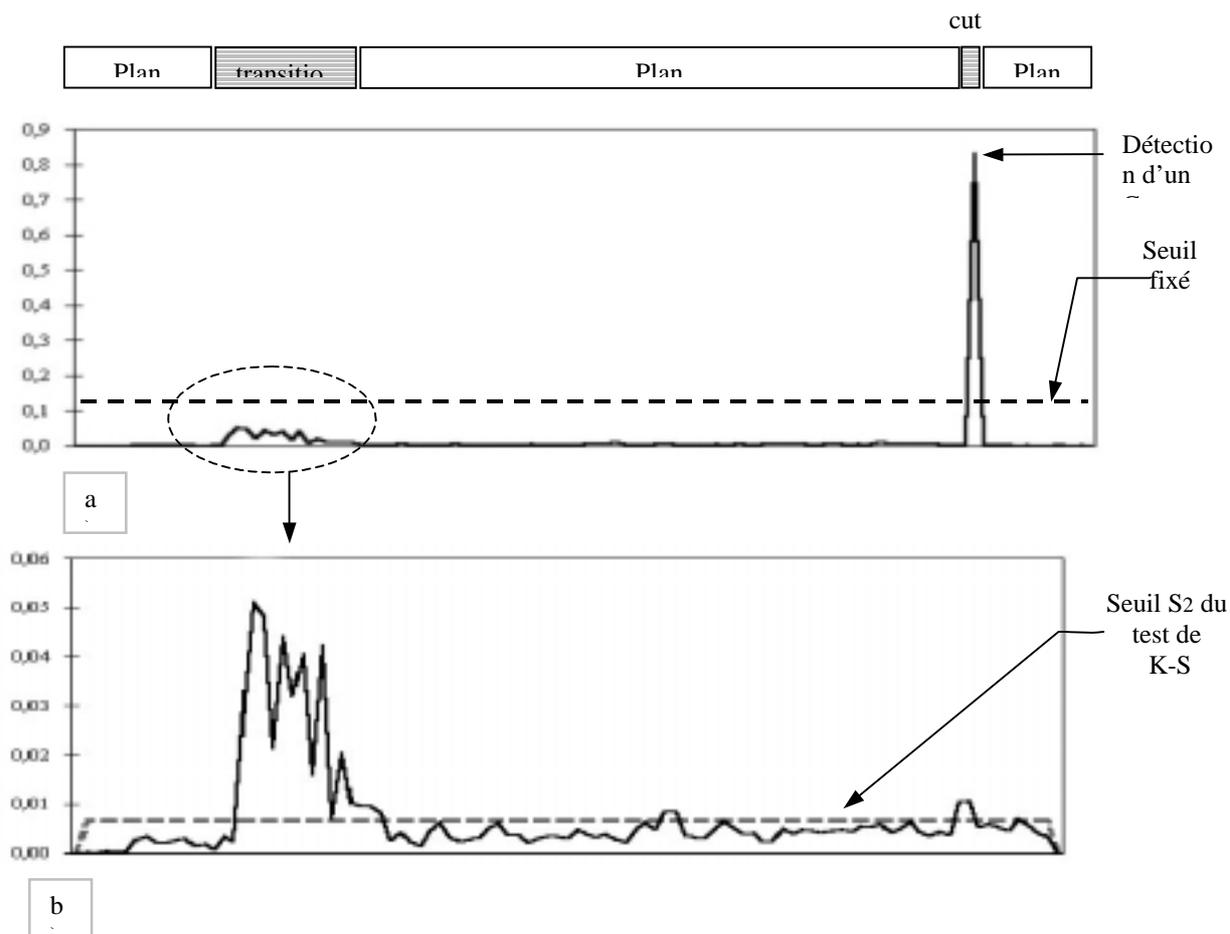


Fig-5 Variation temporelle de la distance de K-S au cours d'une séquence de 150 images. a) sur l'étendu de la séquence on peut vérifier l'existence d'un cut et d'une transition progressive. b) agrandissement du graphe au tours de la transition, les variations modérées mais étendues de la distance de K-S permettent de lancer une détection de transition. On remarque que lors d'une transition progressive la distance de K-S est supérieure au seuil du test de K-S.

6. 2. Paramétrage de l'algorithme

L'algorithme que nous proposons dépend de deux paramètres S_1 et S_2 . Le paramètre S_1 représente le seuil au-dessus duquel la variation de la distance de Kolmogorov-Smirnov (K-S) devient suffisamment grande pour détecter uniquement un changement de plan de type "cut". Ce seuil a été fixé, après plusieurs expérimentations, à la valeur '0.1'. Le seuil S_2

représente la valeur du seuil du test de K-S. Sa valeur dépend de la taille des images (largeur et hauteur) et de la précision désirée du test. Ce seuil est utilisé pour lancer l'opération de détection de transitions progressives. La valeur de ce seuil a été tabulée. La précision du test a été fixée à 1%.

6.3 Résultats expérimentaux

Généralement les séquences utilisées pour valider les algorithmes de détection de changement sont des séquences représentant des scènes réelles dont le contenu est très général. Ce sont des séquences extraites de films ou de clips vidéo, qui doivent contenir les différents types de transitions pour permettre de voir le comportement de l'algorithme vis-à-vis de ces transitions.

Nous allons évaluer l'algorithme de détection de changement de plan proposé sur des séquences MPEG comportant différents types de transitions. La figure 6 présente la détection du changement de plan dans le cas d'un cut. La séquence utilisée est une séquence MPEG que nous avons créée à partir de trois séquences séparées. Deux cuts ont été détectés avec efficacité. La figure 7 présente une détection d'une transition progressive (volet).

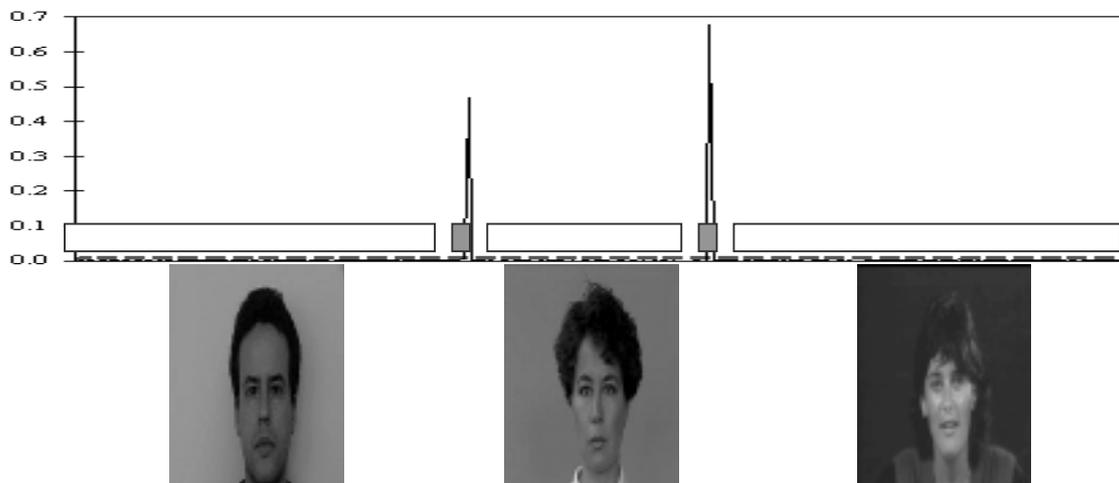


Fig-6 détection d'un cut. Le graphe (en haut) montre l'évolution temporelle de la distance de K-S, le seuil du test de K-S (S_2) est dessiné en pointillé. Les deux pics rencontrés correspondent aux deux cuts détectés. La situation réelle de la position des cuts est montrée par le schéma. En bas : Les images utilisées pour créer la séquence.



a



b



c



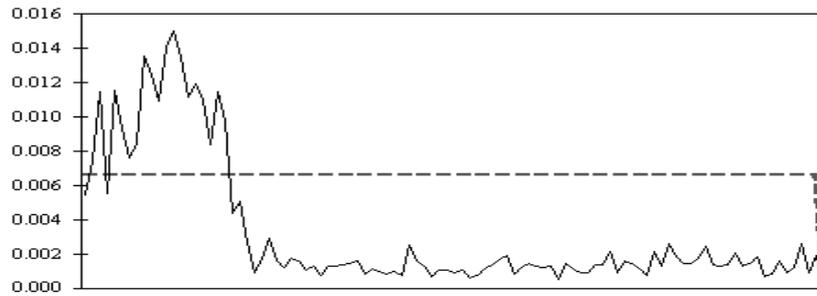
d



e

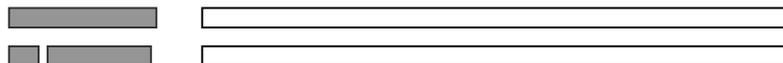


f



g

□ Plan
 ■ Transition



h

Fig-7 *détection d'une transition progressive (volet) dans la séquence " Dinosauré ". a)b)c)d)e)f) image RISTI VOL.11 N° 01 Année 2001 102,104,106,110,120,170. g) évolution temporelle de la distance de K-S. Le seuil du test de K-S (S_2) est dessiné en pointillé. h) situation réelle (dessus) et transitions détectées par l'algorithme proposé (dessous).*

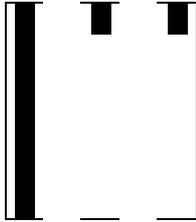
Conclusion

Le découpage temporel de la vidéo en plans est une étape importante de la structuration du contenu de la vidéo. Son importance réside dans le fait qu'elle permet de retrouver les plans qui sont les composants les plus élémentaires manipulés lors de l'opération de montage vidéo. L'approche adoptée consiste à détecter le changement de plans entre les images successives. Elle constitue un compromis idéal entre la complexité algorithmique et la qualité des résultats requise.

La technique que nous proposons est inspirée des méthodes de comparaison d'histogrammes globaux d'intensité. Nous avons essayé de combler les manques de cette méthode qui découlent essentiellement de l'inefficacité des tests statistiques utilisés et de l'utilisation d'un seuil unique qui ne permet pas de détecter les différents types de transitions entre plans. L'utilisation d'un algorithme à deux seuils nous a permis de détecter avec efficacité les différents types de transitions (cuts et transitions progressives). L'introduction du test statistique de Kolmogorov-Smirnov a donné plus de sens à la partition temporelle obtenue qui se rapproche plus du montage réel de la vidéo.

Références Bibliographiques

- [BOU96] P. BOUTHMEY & F. GANANSIA
"Video partitionning and camera motion
characterisation for content-based
indexing" Rapport de recherche, IRISA, 1996.
- [BOU97] P. BOUTHMEY & M. GELGON & R. FABLET
"A unified approach to shot change detection
and camera motion characterisation"
Rapport de recherche, INRIA, 1997.
- [CUE98] A. A. SAVITHA & D. PONCELEON & D. PETKOVIC
"CUE VIDEO : automated video/audio indexing
and browsing"
IBM ALMADEN research center, CA, USA
- [KAS98] R. KASTURI & U. GARGI & S. ANTANI
"Performance characterisation and comparai son
of video indexing algorithms"
Tecnical reports, Pennesylvani a State
uni versi ty, 1998.
- [KEN98] John KENDER & BOON-LOCK YEO
"Video scene segmentation via continous video
coherence"
IEEE, 1998.
- [TAS89] Philippe TASSI
"Méthodes statistiques"
Edition ECONOMICA, 1989.
- [TEN95] P. TENENHAUS
"Programmation multimédia en c sous Windows"
Edition EYROLLES, 1995.



FORMULAIRE CONFIDENTIEL D'EVALUATION D'UN PROJET DE PUBLICATION

N° de code :
Date d'envoi :
Nom du référé :
Intitulé :
.....

EVALUATION DE L'ARTICLE

Mettre une croix dans l'une des cases suivantes :

(A – très bien, B- Bien, C- Assez bien, D- Passable, E- Médiocre)

	A	B	C	D	E
01- Originalité du travail	<input type="checkbox"/>				
02- Pertinence par rapport à la connaissance scientifique dans le domaine	<input type="checkbox"/>				
03- Fondements théoriques et hypothèses.....	<input type="checkbox"/>				
04 – Méthodologie : (description adéquate et appropriée des matériaux et des méthodes).....	<input type="checkbox"/>				
05-Adéquation hypothèses / résultats	<input type="checkbox"/>				
06- Qualité des références bibliographiques..... (75% des références doivent dater de la dernière décennie)	<input type="checkbox"/>				
	<input type="checkbox"/>				

07- Qualité et clarté des tableaux et schémas.....

08- Maîtrise de la langue utilisée.....

09- Style.....

Avis du référé :

(cocher l'une des cases suivantes)

- Accepté dans la forme présentée
- Accepté après révision mineure
- Accepté après révision majeure
- (Assez long. Le manuscrit doit être condensé davantage).
- Rejeté

Signature du référé

Date :

