Détection d’un changement de plan vidéo et son identification

Projet EVASION 2019 : François Beauvois, Nabil Boukraa, Victor Costa et Sofiane Kasmi

# Résumé

(à faire à la fin de la rédaction)

# Introduction

La détection d’un changement de plan vidéo ainsi que son identification est un problème extensivement étudié par la communauté scientifique. En conséquence, l’objectif de l’étude n’est pas de trouver un solution « rupturiste » mais plutôt de mettre en place une démarche scientifique utilisant les outils et méthodes acquises le long du CES en *Big Data*, comme par exemple la classification.

Ainsi, ce rapport décrit la progression pédagogique mis en place pour résoudre ces deux problèmes. L’objectif pratique était de développer des algorithmes sur langage python qui soient les plus génériques et robustes possibles. La méthodologie appliqué passe par la séparation de l’étude en deux parties : 1) la détection de changement de plan et 2) l’identification des plans. Le rapport est donc organisé dans cet optique.

# Sources multimédia et outils

### Sources

La vidéo de référence contient des images en 256x320 pixels, 256 niveaux RGB et une durée de 39m10s. Dans un premier temps, l’étude a été réalisé sur 2351 frames, ce qui correspond à un rythme de 1 frame par minute.

### Indicateurs

Le fichier d’annotations *06-11-22.trs* a été complété (ajout du reportage en début de la vidéo) et utilisé comme Gold Standard pour l’évaluation des méthodes. Les indicateurs utilisées sont les habituelles dans le domaine : nombre de détections, vrai positives, faux positives et faux négatives. Le calcul de la précision et du *recall* viennent compléter ces indicateurs.

### Codage

Le codage a été en langage python avec l’utilisation des bibliothèques habituels (numpy, math…). La bibliothèque utilisé pour le traitement sur image est le *open cv* dans sa version 2.0.

# Détection de changement de plan

La détection d’un changement de plan peut se traduire par la différence sur un ou plusieurs critères entre deux images consécutives. Quand la différence dépasse un ou plusieurs seuils, l’image est retenu.

Deux méthodes sont étudiées pour caractériser une image : les moyennes et les histogrammes. Chaque méthode est décrit dans les sections suivantes.

## Moyennes

La première approche consiste à calculer la moyenne sur un des canaux de couleur et la comparer entre deux images consécutives.

Bien qu’une fonction de lissage a été développé et testé au sein du projet, il a été décidé de ne pas l’utiliser pour maximiser les différences entre deux images consécutives. Cependant, comme l’échantillonnage effectué est assez espacé, le lissage pourrait être utile pour les mouvement de camera (travellings, zooms, …). La technique de lissage entre plans est plus approprié à des images compressées ou qui contiennent une quantité de bruit important.

Le travail sur la décomposition en gris est décrit dans la section suivante. Les canaux restants sont décrits dans les sections postérieurs avec moins de détail.

### Moyenne sur le gris

La moyenne des niveaux de gris a été calculé pour chaque image. La figure suivante présente les niveaux obtenues pour les 500 premiers images.

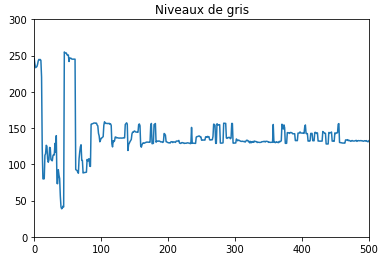


Figure 1 : Diagramme avec la moyenne des niveaux de gris pour les 500 premiers images.

La dérivé par rapport au temps a été ensuite calculé par simple différence entre deux images consécutives (cf. *Figure* 2 à gauche).

|  |  |
| --- | --- |
|  |  |

Figure 2 : Gauche : Diagramme avec la différence entre la moyenne des niveaux de gris consécutifs pour les 500 premiers images. Droite : Histogramme des variations de la moyenne de gris pour l’ensemble des images.

La *Figure* 2 à gauche montre des zones à faible variation intercalés avec des signaux forts qui correspondent à des fortes variations. L’histogramme de la dérivé (*Figure* 2 à droite) montre que la majorité des différences sont relativement petites (entre -3 et 3) mais qui peuvent s’étaler jusqu’à ±30.

Le choix du seuil à définir se fait à l’aide d’une recherche paramétrique (cf. *Figure 3*). Nous avons privilégié le *recall* pour minimiser le nombre de faux négatives. Il est visible

|  |  |
| --- | --- |
|  |  |

Figure 3 : Gauche : Evolution de la précision (bleu) et du recall (orange) avec le seuil de détection (axe des abscisses). Droite : Evolution du nombre de détections avec le seuil de détection.

La *Figure 3* montre également que le nombre de détections décroit rapidement avec l’augmentation du seuil de détection. Le seuil optimal est donc à 3 et permet d’obtenir une précision de 85.4% et un *recall* de 69.7%.

### Moyenne sur les composantes RGB et HSV

La démarche suivi pour le gris a été appliqué a chaque composante de palettes RGB et HSV. Les résultats sont présentés dans le tableau suivante :

Tableau 1 : Indicateurs de performance pour la technique de la moyenne pour différentes composantes de couleur.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Palette* | *Composante* | *Seuil optimal* | *Précision (%)* | *Recall (%)* |
| - | Gris | 3 | 85.4 | 69.7 |
| RGB | Rouge | 3 | 85.8 | 75.7 |
| Vert | 2 | 78.9 | 69.1 |
| Bleu | 3 | 85.4 | 70.2 |
| HSV | Hue | 3 | 85.4 | 70.4 |
| Saturation | 2 | 78.8 | 68.6 |
| Valeur | 3 | 85.8 | 75.6 |

Le Tableau 1 montre que globalement les résultats sont assez proches, avec un meilleur performance pour le traitement effectué avec le rouge et en travaillant avec la luminosité (valeur de HSV). Bien que les résultats puissent entre améliorés en affinant le seuil (réduction du pas de recherche), l’exercice montre la que technique de la moyenne ne permettra pas de dépasser les 80% en *recall*, ce qui n’est pas satisfaisante.

Afin d’améliorer ces résultats, il a été décidé d’utiliser les informations des composantes étudiées en définissant une distance euclidienne avec les 7 canaux utilisées entre deux images. Les résultats sont légèrement meilleurs mais pas encore satisfaisantes (non montré ici). La précision et *recall* ainsi obtenues sont de 85.7% et 78.7%, respectivement. Il est à noter que les niveaux de gris et de HSV étant des projections du RGB, il serai souhaitable d’utiliser uniquement les trois composantes RGB pour ne pas biaiser les distances.

### Séparation de l’image en sections

La technique de séparation en sections permet d’entre plus sensible aux variations locales dans les images. Ainsi, il a été décidé de découper les images en 16 sections uniformes et réaliser l’étude des moyennes par section.

La figure suivante monte les 16 diagrammes pour la dérivé des moyennes des niveaux de gris.

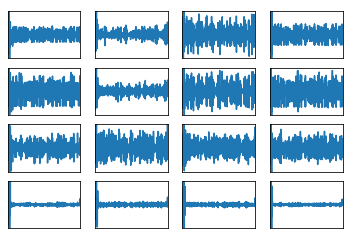


Figure 4 : Diagrammes avec la différence entre la moyenne des niveaux de gris consécutifs pour les 16 sections.

La Figure 4 met en évidence la faible variation des 4 sections inferieurs qui sont utilisées surtout pour les *inserts* graphiques de l’émission.

L’implémentation de cette technique a eu besoin d’un deuxième critère de seuil. Le choix retenu est le nombre de sections d’une image qui ont été identifiées par le premier seuil de différence . L’optimisation des deux seuils a été effectué à partir d’une recherche systématique.

Pour le gris, le meilleur *recall* est à 87.2% pour une précision de 88.1% avec le seuil sur la dérivé à 6.5 et un seuil sur le nombre d’activations à 45% (~7 sur 16). En effet, le découpage de l’image en section a permis d’augmenter d’environ 10 points le *recall* obtenu sans perdre en précision.

Les résultats obtenus pour les autres composantes sont présentés dans le tableau suivante :

Tableau 2 : Indicateurs de performance pour la moyenne par section pour différentes composantes de couleur.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Palette* | *Composante* | *Seuil diff moyenne* | *Seuil nb activations (%)* | *Précision (%)* | *Recall (%)* |
| - | Gris | 6.5 | 45 | 88.1 | 88.1 |
| RGB | Rouge | 7 | 50 | 89.9 | 88.0 |
| Vert | 8 | 50 | 90.2 | 87.4 |
| Bleu | 10 | 50 | 89.9 | 87.7 |
| HSV | Hue | 10 | 50 | 90.1 | 88.2 |
| Saturation | 8 | 50 | 90.2 | 87.5 |
| Valeur | 7 | 50 | 89.9 | 88.0 |

Comme pour les moyennes, la combinaison des 7 composantes a permis d’obtenir une précision de 90.6% et un *recall* de 89%. Globalement, les résultats sont relativement proches entre eux.

## Histogrammes

L’utilisation des moyennes pour caractériser une image ou une sous-section est un moyen rapide d’obtenir un descripteur mais peu s’avéré insuffisante quand la précision est recherché ou quand plus de descripteurs sont nécessaires pour faire de la classification.

Ainsi, l’utilisation des histogrammes des niveaux d’une ou plusieurs composantes de palette chromique est le choix d’élection.

Dans le présent étude, seul l’histogramme de niveaux de gris a été utilisé en combinaison avec la décomposition de l’image en 16 sections. Il a été décidé de travailler avec des histogrammes à 64 niveaux (bins). De plus, les histogrammes ont été normalisées en utilisant la norme L1 pour éviter la création de biais entre les composantes de couleur. La norme L2 a été également testé avec des résultats moins bons.

Une fois les histogrammes normalisés, la distance χ2 a été calculé entre images consécutives. Cette distance est la plus adapté à comparer des histogrammes. Le formule utilisé est la suivante :

|  |  |
| --- | --- |
|  | Équation 1 |

Cette déclinaison de la distribution χ2 est disponible dans la bibliothèque python *open cv* sous le nom de *HISTCMP\_CHISQR\_ALT*. L’avantage est que la distance peut être calculé entre deux histogrammes qui ont des valeurs nuls, comme les images en blanc. La division par zéro est ainsi évité si une des images est non nul.

La figure suivante présente un exemple d’un diagramme de distances d’une section pour les 100 premières images de l’étude :

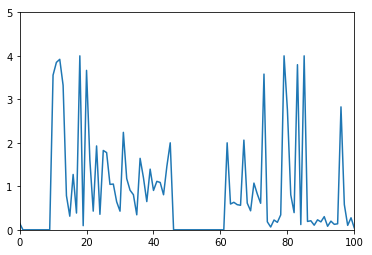


Figure 5 : Diagrammes avec la distance χ2 dans la section 1 pour les 100 premières images.

La Figure 5 montre des pics d’intensité très bien définis. De plus, les zones identiques sont visibles jusqu’à l’image 10 et entre l’image 45 et 60.

Comme précédemment, deux seuils ont été mises en place pour identifier un changement de plan : seuil sur la distance χ2 et seuil sur le nombre de sections activés. L’optimisation a été effectué par recherche systématique. Les résultats sont présentés dans le tableau suivant :

Tableau 3 : Configuration et performances pour la distance χ2 sur 64 bins de la composante gris sur 16 sections.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Composante* | *Seuil distance* | *Seuil nb activations (%)* | *Précision (%)* | *Recall (%)* | *Nb. plans « vrai »* | *Nb. Plans identifiées* | *Vrai positives* | *Faux positives* | *Faux négatives* |
| Gris | 1.5 | 65 | 95.3 | 95.3 | 512 \* | 536 | 511 | 25 | 0 |

*\* Il existe deux changement de plans dans la même minute et ont donc la même image comme référence, d’où la différence entre le nombre de plans « vrai » et le nombre de plans identifiées.*

Le Tableau 3 montre une amélioration significative de cette méthode par rapport aux techniques basées sur les moyennes. Il est surtout à retenir la non existence de faux négatives. Il est donc recommandé d’utiliser une deuxième technique pour l’élimination des 25 faux positives.

## Conclusion partiel

# Identification des plans

# Conclusion