

Génération de séquence résumé par une nouvelle approche basée sur le Soft Computing

Youssef Hadi^{*,**}, Rachid El Meziane^{***} Rachid Oulad Haj Thami^{*,**}

^{*}Laboratoire GSCM_LRIT, Faculté des Sciences,
Université Mohamed V Agdal B.P. 1014 Rabat, Maroc
hadiyoussef@gmail.com,

^{**}Laboratoire SI2M, Equipe WiM, ENSIAS,
Université Mohamed V Souissi B.P. 713 Rabat-Agdal, Maroc
oulad@ensias.ma

^{***}ENSIAS, Université Mohamed V Souissi,
B.P. 713 Rabat-Agdal, Maroc
meziane@ensias.ma

Résumé. Cet article propose une approche d'abstraction des séquences vidéo basée sur le soft computing. Etant donné une longueur cible du condensé vidéo, on cherche les segments vidéo qui couvrent le maximum du visuel de la vidéo originale en respectant la longueur du condensé.

1 La génération des séquences résumées proposée

La génération de résumé vidéo est une technique alternative prometteuse utilisée dans l'indexation et la recherche vidéo (Truong et Venkatesh, 2007). L'objectif de cet article est de proposer une approche de génération des séquences résumées utilisant le soft computing vue son efficacité dans les systèmes tolérant l'imprécision et l'incertitude (Zadeh, 1956). Etant donné une longueur cible du condensé vidéo, on veut calculer les segments vidéo qui couvrent le maximum du visuel en respectant la longueur du condensé. Les segments vidéo sont représentés par des images clés et décrites par un histogramme de couleur. L'histogramme est un outil très ordinaire pour résumer visuellement la distribution d'un échantillon de données. Pour maximiser le visuel on doit maximiser la présence de l'information de la couleur et sa distribution. Ce pendant, le problème revient à maximiser l'*occurrence* et la *distribution* du visuel représenté par l'information de couleur du contenu vidéo. En effet, l'*occurrence* et la *distribution* de ce contenu visuel représente des variables linguistiques, c.-à-d. "l'*occurrence* est maximale" et "la *distribution* est maximale". L'*occurrence* et la *distribution* représentent deux ensembles flous qui nécessitent une détermination de leur fonction d'appartenance.

Dans notre proposition, nous segmentons les séquences vidéo en plans (*shots*) comme unité de base en utilisant une transformation de couleur *RGB* réversible comme une représentation du contenu vidéo (Hadi et al., 2006b). Après, nous sélectionnons les images représentatives (*keyframes*) à partir des plans vidéo on se basant sur l'estimation du mouvement local (Essannouni et al., 2006). L'algorithme d'extraction des images représentatives utilisé est basé sur un

algorithme de classification k-medoid pour rechercher les images les plus représentatives pour chaque plan Hadi et al. (2006a). Comme résultat, les images extraites représentent le centre des sous segments dans un plan.

Dans l'optique de créer des condensés vidéo à partir des segments résultats, nous représentons les images clés calculées par leur histogramme de couleur utilisé dans (Hadi et al., 2006b). La création est réalisée en maximisant l'*occurrence* et la *distribution* de la couleur considérées comme des ensembles flous. Afin de déterminer leurs fonctions d'appartenance, on cherche les instances d'histogrammes liées aux *occurrences* maximales et aux *distributions* maximales en utilisant les algorithmes génétiques (AG). Avec ces deux instances fictives d'histogrammes, nous calculons leurs distances avec les histogrammes des images clés calculées dans la première étape d'analyse vidéo. On obtient ainsi deux vecteurs de distance pour l'ensemble des segments à sélectionner pour la création du condensé vidéo. Afin de chercher la valeur d'appartenance du visuel maximal, nous utilisons les théories des ensembles flous. Le condensé final sera généré à partir de la courbe des valeurs d'appartenance du visuel maximal en sélectionnant les segments ayant les plus grandes valeurs (*visuel maximal*) en respectant la longueur du condensé qui représente la somme des longueurs des segments sélectionnés.

Références

- Essannouni, F., Y. Hadi, R. Oulad Haj Thami, D. Aboutajdine, et A. Salam (2006). A new optimal frequency motion estimation algorithm. In *Proceedings of the second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, Marrakech, Morocco.
- Hadi, Y., F. Essannouni, et R. Oulad Haj Thami (2006a). Video summarization by k-medoid clustering. In *SAC '06 : Proceedings of the 2006 ACM symposium on Applied computing*, New York, NY, USA, pp. 1400–1401. ACM Press.
- Hadi, Y., F. Essannouni, R. Oulad Haj Thami, et D. Aboutajdine (2006b). A new approach for video cut detection using color histogram. In *Proceedings of the 3rd International Symposium on Image/Video Communications over fixed and mobile networks, ISIVC'2006*, Hammamet, Tunisia.
- Nuñez-García, J. et O. Wolkenhauer (2001). Random sets and histograms. In *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, Volume 3, pp. 1183–1186.
- Truong, B. T. et S. Venkatesh (2007). Video abstraction : A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.* 3(1), 3.
- Zadeh, L. A. (1956). Soft computing and fuzzy logic. In *IEEE Software*, Volume 11, pp. 48–56.

Summary

This paper propose a video skimming approach by using soft computing. Given a targets length of the video skim, we look for video segments that can cover the maximum of the visual content while respecting the skim length.