

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