

Prédiction de séries temporelles et applications à l'analyse de séquences vidéo

Rémi Auguste, Ahmed El Ghini, Marius Bilasco, Chabane Djeraba

LIFL/CNRS-UMR 8022 - Université Lille 1

{remi.auguste, ahmed.elghini, marius.bilasco, chabane.djeraba} @lifl.fr

L'utilisation de méthodes de classification avec différentes mesures de similarité a été fortement étudiée pour l'analyse de séquences vidéo (Veeraraghavan et al. (2005), Chen et Zhang (2008) et Simovici et Djeraba (2008)). Ces mesures de similarité appliquées aux séquences d'images sont fondamentales pour la mise en place de systèmes de recherche et d'exploration de vidéos par le contenu visuel. L'approche stochastique basée sur des modèles spatio-temporels relève d'une importance cruciale car elle tient compte des propriétés à la fois spatiales et temporelles de la séquence d'images. Dans cet article, nous étudions le rapprochement entre des séquences vidéo à l'aide de distances entre modèles paramétriques.

Nous utilisons le modèle paramétrique $ARMA(p, q)$ pour modéliser l'évolution des pixels dans le temps, qui peut être considérée comme une série temporelle. Dans la littérature, plusieurs méthodes, plus ou moins complexes, permettent de mesurer la similitude entre deux modèles. Nous proposons d'utiliser la distance euclidienne mentionnée dans l'article de Sekita et al. (1992), généralisée aux modèles $ARMA(p, q)$. Nous comparons cette distance avec une distance plus complexe : la distance de Martin (Martin (2000)) utilisée dans l'article de Chen et Zhang (2008). La distance de Martin est moins précise mais plus robuste que la distance euclidienne pour comparer deux modèles. Elle offre en plus une certaine souplesse, ce qui s'avère utile lorsque les vidéos modélisées sont sujettes au bruit ou à des conditions d'éclairage différentes. En revanche, le coût de calcul de la distance euclidienne est moindre ($O(n)$) comparé à celui de la distance de Martin ($O(n^4)$), cette différence peut être particulièrement utile dans le cadre d'une application temps réel. Notre idée est de combiner ces deux distances de façon à obtenir une mesure de similarité qui concilie à la fois vitesse, simplicité et précision. Nous calculons tout d'abord la distance euclidienne, si cette mesure montre que les deux modèles sont semblables alors une mesure plus précise est calculée grâce à la distance de Martin. Fixer les seuils pour considérer deux modèles semblables ou non reste un inconvénient majeur quelque soit la distance utilisée. Une piste possible est de se baser sur la moyenne des paramètres des modèles pour déterminer ces seuils.

Nous allons maintenant nous intéresser à la façon d'exploiter les propositions précédentes dans la conception d'une application de reconnaissance d'objet ou de personne. Voici les différentes étapes que nous proposons pour construire une telle application :

Segmentation de l'image La modélisation complète de l'image n'offre pas de bonnes performances ni de données exploitables, la solution souvent employée est de découper l'image en sous images (segments) qui sont modélisées séparément. Une des difficultés rencontrée est de dimensionner correctement ces sous-images pour obtenir un bon rapport précision/performance.

Modélisation des segments La couleur moyenne de chaque segment est modélisée sur l'ensemble des images de la vidéo. Chaque séquence vidéo obtient ainsi un ensemble de modèles correspondant aux segments.

Reconnaissance de l'image Afin de reconnaître une séquence vidéo nouvellement acquise, il est nécessaire de la modéliser en utilisant les étapes précédentes. Ensuite, chaque modèle d'une vidéo est comparé avec tous ceux d'une autre. Cette comparaison est réalisée grâce à notre mélange de distance. Pour chaque modèle, nous conservons uniquement la plus petite distance mesurée. Nous effectuons ensuite la moyenne de ces distances. Si cette moyenne est faible, par rapport à la valeur moyenne des coefficients de tous les modèles alors ces deux séquences vidéo peuvent être considérées comme similaires.

Nous travaillons actuellement sur l'implantation d'une adaptation de cet algorithme à des descripteurs de mouvement (flux optique) pour reconnaître des actions. Cette implantation nous permettra d'évaluer les performances de notre algorithme. Nos travaux futurs sont basés sur des modèles multidimensionnels qui nous apportent plus d'information au niveau l'évolution temporelle des descripteurs et la caractérisation du contenu dynamique de la séquence (cf Ma et al. (2009) et Wang et al. (2008)); car ils tiennent compte de l'interdépendance entre les différentes composantes temporelles de ces descripteurs.

Références

- Chen, W. et Y. Zhang (2008). Parametric model for video content analysis. *Pattern Recognition Letters* 29, 181–191.
- Ma, Y., P. Cisar, et A. Kembhavi (2009). Motion segmentation and activity representation in crowds. *2009 Wiley Periodicals, Inc. Int J Imaging Syst Technol* 19, 80–90.
- Martin, R. (2000). A metric for ARMA processes. *IEEE Trans. on Signal Processing* 48(4), 1164–1170.
- Sekita, I., T. Kurita, et N. Otsu (1992). Complex autoregressive model for shape recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(4), 489–496.
- Simovici, D. et C. Djeraba (2008). *Mathematical Tools for Data Mining*. Springer. Set Theory, Partial Orders, Combinatorics.
- Veeraraghavan, A., A. Roy-Chowdhury, et R. Chellappa (2005). Matching shape sequences in video with an application to human movement analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27(12), 1396–1909.
- Wang, L., X. Wang, C. Leckie, et K. Ramamohanarao (2008). *Advances in Knowledge Discovery and Data Mining*, Chapter Characteristic-Based Descriptors for Motion Sequence Recognition, pp. 369–380. Springer.

Summary

In this article, we present a statistical method which allows us to measure the similarity between videos, it is based on a generalized distance between parametric models. The final purpose of our work is to recognize objects or people.