

Classification des actions humaines basée sur les descripteurs spatio-temporels

Sameh Megrhi, Azeddine Beghdadi, Wided Souidène

L2TI, Institut Galilée, Université Paris 13, Sorbonne Paris Cité,
avenue J. B. Clément 93430 Villetaneuse, France

Résumé. Dans cet article, nous proposons un nouveau descripteur spatio-temporel appelé ST-SURF pour l'analyse et la reconnaissance d'actions dans des flux vidéo. L'idée principale est d'enrichir le descripteur *Speed Up Robust Feature* (SURF) en intégrant l'information de mouvement issue du flot optique. Seuls les points d'intérêts qui ont subi un déplacement sont pris en compte pour générer un dictionnaire de mots visuels (DMV) robuste basé sur l'algorithme des k-moyennes (K-means). Le dictionnaire est utilisé lors du processus d'apprentissage et de reconnaissance d'actions basé sur la méthode des machines à vecteurs supports (SVM). Les résultats obtenus confirment l'intérêt du descripteur proposé ST-SURF pour l'analyse de scènes et en particulier pour la reconnaissance d'actions. La méthode atteint une précision de reconnaissance de l'ordre de 80.7%, équivalente aux performances des descripteurs spatio-temporels de l'état de l'art.

1 Introduction

Le stockage et la distribution à travers les média numériques, et plus particulièrement internet, de données visuelles atteignent des proportions gigantesques. Ceci est accéléré par la banalisation des outils de capture et d'éditations de données numériques telles que la vidéo et l'audio. Cette masse de données de différentes modalités représente une information capitale pour les étapes d'identification d'événements et de décision. Il est donc nécessaire de développer des solutions automatiques pour analyser ces contenus numériques. Une des problématiques qui suscitent beaucoup d'intérêt depuis quelques années est la détection et la reconnaissance des actions humaines dans les séquences vidéo. On appelle action tout événement caractérisé par des mouvements ou de comportements anormaux que l'on rencontre par exemple dans les flux de vidéo surveillance Bouttefroy et al. (2010). La détection d'actions trouve de nombreuses applications telles que l'indexation, la vidéo surveillance Hu et al. (2004) ou le résumé de vidéos Zhou et al. (2008), pour ne citer que quelques-unes. Dans le contexte de la reconnaissance de l'action, la représentation des descripteurs vidéo au moyen d'un dictionnaire de mots visuels, est un domaine de recherche très actif, Willamowski et al. (2004). L'idée de base d'un DMV est de grouper un ensemble d'objets, par exemple des descripteurs visuels, en groupes de sorte que les objets de même type soient dans un même groupe (cluster). Récemment, l'algorithme des K-moyennes a été largement utilisé pour construire les DMV en