

Extraction et reconnaissance des objets dans des scènes vidéo complexes : application au monitoring vidéo des patients atteints de la maladie Alzheimer

Lieu : LABRI UMR 58 00/ Université Bordeaux 1,  
Equipe : Image et Son, thème Analyse et Indexation Vidéo.

La détection automatique et reconnaissance des objets dans la vidéo est le problème clé traité à ce jour par la vaste communauté internationale [1]. L'intégration de la modélisation de perception visuelle des scènes dans la méthodologie d'extraction des objets préalable à leur reconnaissance est une des directions prometteuses. Ainsi la définition de la saillance visuelle des parties de la scène permet de localiser les zones d'intérêt où les objets doivent être recherchés [2,3]. Les travaux actuels en définition de la saillance visuelle des scènes vidéos menés au LABRI [4] serviront de base pour une extraction efficace et automatique des zones d'intérêt à fin de limiter les zones de recherche d'objets.

Ensuite se pose le problème de la description et de reconnaissance des objets tenant compte d'une très grande variabilité des objets de même catégorie. Dans nos scènes en provenance de caméras embarquées sur des patients ou encore filmés par une caméra stationnaire la reconnaissance d'une cafetière par exemple dans les vidéos enregistrées au domicile des sujets est un problème non contraint vu la grande variabilité d'apparence et diversité des points de vue due aux mouvements propres des sujets. Les méthodes d'état de l'art en indexation des objets s'appuient sur des descripteurs locaux (points d'intérêt SIFT, SURF...). Pour ces derniers, la signature de la vidéo ou des objets dans la vidéo est construite à partir des caractéristiques du voisinage local des points. Cependant, ces descripteurs sont peu répétables et ne tiennent pas compte de la répartition spatiale des points d'intérêt. Une des solutions pour la prise en compte du contexte spatial devenu populaire consiste à partitionner le plan image en blocks arbitraires et de décrire l'ensemble des singularités locales dans chaque bloc [5]. Nous avons proposé une autre approche : les points d'intérêt sont « connectés » localement dans de petits graphes [8] qui conservent des bonnes propriétés d'invariance par rapport aux transformations affines de l'image : translation, zoom, rotation. La construction des dictionnaires visuels, quantification des graphes en « GraphWords » et la comparaison des signatures des objets ainsi définies constitueront le deuxième volet de la thèse. Intégration de la saillance visuelle dans la construction des petits graphes et leur comparaison dans l'espace de description doivent rendre cette description plus robuste et raisonnablement redondante. Cette thèse s'inscrit dans le contexte de la recherche pionnière menée au LABRI sur interprétation des flux acquis avec les caméras portée par des patients (maladie d'Alzheimer)[6],[7]

<http://immed.labri.fr/>.

Le thésard effectuera son travail de recherche dans le contexte international et pluridisciplinaire du projet européen FP7 Dem@care, il participera aux réunions du projet à l'international

Contact : Jenny Benois-Pineau : [jenny.benois@labri.fr](mailto:jenny.benois@labri.fr)

1. B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In CVPR, 2010.
2. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In CVPR, 2009.
3. B. Alexe, Th. Deselaers and V. Ferrari, "Measuring the objectness of image windows", ETH Zurich, D-ITET, BIWI Technical Report No 276
4. H. Boujut, J. Benois-Pineau, T. Ahmed, O. Hadar, P. Bonnet: A metric for no-reference video quality assessment for HD TV delivery based on saliency maps. ICME 2011: 1-5
5. S. Lazebnik, C. Schmid, J. Ponce, "Beyond the bags of features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006
6. R. Mègret, V. Dovgalecs, H. Wannous, S. Karaman, J. Benois-Pineau, E. El Khoury, J. Pinquier, P. Joly, R. André-Obrecht, Y. Gaëstel et al "The IMMED Project: Wearable Video Monitoring of People with Age Dementia" ACM Multimedia 2010, Octobre 2010, Florence, Italie
7. S. Karaman, J. Benois-Pineau, R. Mègret, V. Dovgalecs, J.-F. Dartigues, Y. Gaëstel ICPR 2010, Août 2010, Istanbul, Turquie
8. S. Karaman, J. Benois-Pineau, R. Mègret, A. Bugeau, "Multi-Layer Graph Words for Object Recognition". 18th International Conference on MultiMedia Modeling (MMM'2012), 4-6 Jan, 2012, Klagenfurt, Autriche.