STARS group works on video analytics. However there are scientific challenges in people monitoring when dealing with real word scenes: cluttered scenes, handling wrong and incomplete person segmentation, handling static and dynamic occlusions, low contrasted objects, moving contextual objects (e.g. chairs), viewpoints...

Several investigations have been carried out to model human activities. For instance, Deep Convolutional Neural Network CNN algorithms [1, 4] have been applied with great success to videos, related to monitoring applications. However, existing work has either focused on simple activities in real-life scenarios, or the recognition of more complex (in terms of visual variabilities) activities in hand-clipped videos from a frontal viewpoint. We still lack research on generic methods that can recognize complex activities in a long video recorded in the wild, from any viewpoint.

Objectives

This work consists in proposing novel Deep Learning algorithms for view invariant activity recognition in videos, which are efficient in real-world settings, where only small training datasets are available (or using only generic training datasets recorded for different sites) and can dynamically adapt to changing environment.

The performance of the current system strongly relies on the camera angle used: provided that the camera angle used in testing is the same (or close to) as the camera angle used in training, the system performs well. On the contrary, the performance drops when a different camera is used. In this work package, we aim at improving the performance by using 3D human pose information. For the extraction of the pose information CMU’s openpose, but any other more performing software can be used. To generate extra views, Generative Adversarial Network (GAN) [2] can be used together with the human pose information to complete the training dataset. Another promising solution would be to train a GAN using sparse representations of human poses [3]. Then, the newly generated poses can be transformed to human images by means of a pose transfer algorithm. The validation dataset will include NTU RGB+D and smart-home [4] datasets.
Bibliography

3. A. Ben Tanfous, H. Drira, B. Ben Amor. Coding Kendall’s Shape Trajectories for 3D Action Recognition. In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018