

Dynamic Decomposition of Scenes

Dave Russell, Vision Group

According to Wikipedia it is estimated that there are 400,000 video surveillance cameras in London alone, and 4 million spread throughout the UK. Considering that these devices operate continuously, the quantity of video data generated is enormous. Both metropolitan authorities and private security firms employ people to watch and monitor ongoing events in the hope of identifying criminal activity, untoward behaviour, and serious but non-malicious situations. Evidently, due to the sheer volume of data, the task of such surveillance becomes increasingly difficult even in a well staffed camera control room. Most humans can maintain the required level of concentration for as little as 20 minutes at a time before fatigue and boredom erode their powers of attention, this being partly due to the fact that for almost all of that time, nothing of interest actually happens. However, thanks to techniques in the field of computer vision, it may now be possible to relegate the bulk of this thankless task to machines which never tire or lose interest or job satisfaction.

By employing machine learning algorithms, activity in a scene covered by a fixed view camera may be accumulated in a *statistical* sense, such that any unusual event is highlighted and brought rapidly to the attention of a human observer. Thus the overall monitoring task can be accomplished with less people watching more stimulating activity for a greater proportion of the time, and in a smaller control room consisting of less TV screens. Instead, the monotonous 'watching' process goes on in the CPU of a computer, which is continuously fed with digital video data from one or more cameras.

But the implementation of software and algorithms suitable for this purpose is a non-trivial task. Machine learning has been around for many decades, and is a diverse topic which has facilitated numerous applications which impinge on many aspects of our daily lives in this modern technological age.

The purpose of this talk is to introduce the concept of *statistical scene analysis* from an algorithmic perspective, with a view to detection of rare events. In general, activity implies movement of objects in a scene such as people, vehicles and trees causing *localized* changes in the scene's appearance due to occlusion of the background. At the same time, variations in prevailing weather and lighting conditions cause changes in scene appearance on a *global* scale. In both cases, the activity manifests itself as changes in colour and intensity at every pixel. Thus the objective is to identify unusual events by modelling the *spatial* and *temporal* intensity characteristics of the scene on a frame by frame basis. The ideal model is constructed or 'learned' incrementally from such data, and matches all aspects of normal activity. Rare behaviour in the scene is highlighted where it occurs because it *does not* fit the model.

The idea of a stochastic model of normal background scene activity is well established in techniques involving Gaussian Mixture Models and Principal Component Analysis. However, it is envisaged that a more elaborate approach whereby a scene is decomposed into *regions* and *layers* characterized by the *nature of their activity content* will ultimately be more successful.