**Abstract**

One of the main targets of computer vision is to interpret the content of image and video data. To interpret image content, one of the essential goals is to build a model depending on a known set of features extracted from image and video data. The built model is then employed to produce inference of the unknown image or video. Many of the latest spectacular achievements of machine learning and deep learning, its dominant existence in many industries, and its fast penetration in our daily work and life are approving that, these techniques are extremely useful for generating the model. In this talk, computer vision techniques with the help of machine learning and deep learning are introduced, including: 1) Object detection in visual surveillance environment 2) Visual object tracking in different challenging scenarios like occlusion, illumination variation etc. 3) Classification/recognition of different types of objects and situations like automatic detection of violence and non-violence area.