# PREFACE

Video surveillance is an active area of research because of its various applications in human robot interaction, entertainment, education, training, video conferencing, and human behavior analysis. Recognition of activities provides important cues for human behavior analysis techniques. In this thesis, the problems of human activity recognition(HAR) in video surveillance systems have been addressed, and different models using conventional machine learning and deep learning has been proposed. Even though this is a well-explored problem in the field of computer vision, many challenges still remain when one is presented with realistic data. These challenges include highly dynamic background, gradual and sudden illumination changes, camera jitter, shadows, reflections that can provoke false detections, waves on the water surface, boat wakes, and weather issues.

First of all, in this thesis, a detailed literature survey has been presented, which includes a survey of various approaches for Human Activity Recognition and study on the evolution of modern datasets for human activity recognition.

Survey of human activity recognition approaches have been further accomplished in two parts - conventional/ hand-crafted features based approaches and deep learning-based approaches. Moreover, the hierarchy of different approaches under them have been discussed, and research gaps have been identified. Further, we have done a comprehensive literature survey on the evolution of modern datasets for human activity recognition. With the evolution of the human activity recognition field, the datasets used for training and testing of the proposed models have also undergone considerable change. In our survey, we attempted to classify and describe a verity of datasets for researchers to choose the

most suitable benchmark for their domain. A set of characteristics have been proposed by which datasets may be compared. Finally, a detailed list of the dataset used for training and evaluation of proposed models have been presented, followed by a discussion on performance measures used in this thesis.

The first model proposed in this thesis for human activity recognition is based on handcrafted features using a hidden markov model (HMM). The proposed model proceeds in three major steps: First, object is segmented using background estimation and subtraction. In second step, feature is extracted and a combination of contour-based distance signal, optical flow and uniform LBP has been generated as feature. Finally, the activity is referenced using a set of Hidden Markov Model. Successfully tested the proposed model on our own view point dataset, KTH Dataset, MSR view point dataset, i3DPost multi-view dataset and WVU Dataset. The experimental results and analysis done shows that the proposed framework works well with respect to multiple view activity recognition for restricted environment datasets and low-level activities.

Second, we have used deep learning based approaches for HAR using automatic feature extraction and classification. In this direction, first model that was proposed, used depth map sequence generated from RGB-D sensors. The proposed model was trained using raw depth sequences of NTU-RGB +D, MSRAction3D and MSRDailyActivity3D datasets considering its capabilities to record geometric information of the object and an enlarged time dimension convolution has been applied for training the model using spatio-temporal features. Since depth is more discriminatory and insensitive to lighting changes in comparison to RGB videos, it is very suitable for activity recognition. Further, by using raw depth data, we are also saving the preprocessing time. We have experimented the impact of larger

spatial resolution and observed that accuracy stabilizes at larger spatial sizes. Experimental results also demonstrate that by lengthening the temporal resolution, we achieve significant improvement in the accuracy. The proposed framework was evaluated on three depth datasets NTU-RGB +D, MSRAction3D and MSRDailyActivity3D. It can be observed that the result obtained is comparable to state of the art models proposed in the literature.

Further, we have extended our research by presenting two deep learning based model using deep residual networks. In first model, we presented a two-stream framework for human activity recognition using spatial and spatio-temporal features as two streams of the model. Here, we investigated and used deep residual networks with decision fusion based dual stream model for activity recognition from video streams. The architecture is trained and evaluated using standard video actions benchmarks of UCF-101, HMDB-51 and NTU RGB. Performance of depth-based variants of residual networks is also analyzed. The proposed approach not only provides competitive results but also better at exploiting pre-trained model and annotated image data. Second model using residual networks is an encoder-decoder based technique using CRNN, which is combined implementation of CNN as encoder and RNN as a decoder. Introduction of residual connections in traditional CNN model to design very deep architectures known as residual networks are very efficient for computer vision tasks. To exploit capabilities of both CNN and RNN, the proposed model is based on CRNN, which is trained from scratch as well as using ResNet 152, which is pre-trained on ImageNet dataset. The architecture is trained and validated on popular UCF-101 dataset on the basis of accuracy and average loss. From the obtained results, it can be observed that the proposed approach provides better results than many state of art methods.

Finally, a four-stream model combining CNN streams of dynamic images and depth map data has been proposed. The proposed model uses a deep learning model for recognizing human activities in a video sequence by combining multiple CNN streams. The proposed work comprises the use of dynamic images generated from RGB images and depth motion maps for three different dimensions. The proposed model is trained using these four streams on VGG network for action recognition purpose. Further, it is evaluated and compared with the other state of the art methods available in the literature, on three challenging datasets namely MSR daily Activity, UTD MHAD and CAD 60, in terms of accuracy, error, recall, specificity, precision and f-score. From obtained results, it has been observed that the proposed method outperforms other methods.