Chapter 7 CONCLUSION AND FUTURE WORK

The conclusions of the work in this thesis and suggestions for future research are presented in this chapter.

7.1. Conclusions

Enabling computers to understand images and videos of real scenario is a crucial step towards the development of any intelligent video surveillance system. Intelligent video surveillance has vast applications which includes transport applications (such as monitoring of railway station, airport, traffic control etc.), industrial applications (such as monitoring different industry plants), and security applications (such as monitoring of indoor and outdoor environments, monitoring of people in different places etc.).

Focus of this thesis is to develop some effective algorithms for different applications of intelligent video surveillance. The developed methods should be robust as much as possible with appearance of different objects and in complex environments. In this thesis, novel methods using traditional and Deep Learning approaches for human activity recognition are proposed followed by their testing and evaluation as compared to other representative state-of-the-art methods.

Chapter 2 discusses the theoretical background for video surveillance system. In this chapter, we have also given an overview of deep learning and machine learning approaches. Further, in this chapter a literature survey of prominent approaches for HAR using conventional and deep learning approaches are given. Furthermore, it presents a detailed survey along with evolution of modern datasets for HAR.

Chapter 3 presents a HAR method based on conventional machine learning approach. Complete framework comprises of three steps. In first step, we perform

preprocessing and background subtraction. Secondly, feature computation is done. In third and last step, we have used HMM for activity modeling and classification.

Chapter 4 presents a Deep Learning based HAR approach utilizing depth map sequences as input. The proposed approach uses enlarged temporal dimension of depth map sequences as input for training the deep neural architecture. The impact of enlarged larger temporal and spatial resolution has been evaluated on three HAR depth datasets namely NTU-RGB +D, MSRAction3D and MSRDailyActivity3D. From the experimental results, it can be observed that the result obtained is comparable to state of the arts models proposed in literature.

Chapter 5 of the thesis presents HAR models utilizing deep residual neural networks. The first model is dual stream model using residual-CNN of two streams namely spatial and spatio-temporal. While the other discussed model is encoder-decoder based model using CRNN. Which is combination of CNN as encoder and RNN as decoder.

Chapter 6 presents a HAR model by a combination of different modalities from RGB-D sensor. In this work, dynamic images trained on pre-trained VGG-F network and depth images for different views such as top, side and front separately trained on pre-trained VGG-F network are combined. We have tested the network on most promising datasets such as MSR Daily Activity, UTD MHAD and CAD 60 and achieved state-of-the-art results. In addition, we have compared our results for different datasets and found that the proposed method outperforms most of the available methods.

Finally, the overall conclusion of the thesis is being summarized as follows:

- Performed extensive study of the existing literature on experiments and research performed under Human Activity Recognition using conventional as well deep learning approaches to identify research gaps.

- Performed extensive study of literature on evolution of modern datasets for human activity recognition.
- Proposed a HAR framework using HMM to recognize human activities in controlled/ lab environment.
- Proposed and evaluated performance of enlarged temporal convolution network on depth sequences received from RGB-D Sensors. Further, analyzed performance of the model for varying temporal dimension.
- Proposed an encoder decoder based framework using ResNet CRNN. Evaluation of the proposed model by training the network from scratch and by using pre-trained ResNet.
- Proposed a two-stream model using spatial and spatio-temporal streams for activity recognition from videos. Implemented the proposed model and fine-tune pre-trained model for proposed solution. Further, analyzed the impact of network depth on the performance of proposed model.
- Proposed a four-stream model using depth and RGB images generated from RGB-D Sensors. Dynamic images from RGB input using rank pooling were generated and used as the first stream. Further, depth motion maps in three different directions namely front, side and top were generated and used as three different streams to the model. Individual CNN streams were trained using these 4 inputs. Finally, using weighted product model, decision fusion was done for the outputs received from each stream.

7.2. Suggestions for Future Research

The research work presented in this thesis can be taken further into different directions. The scope for future works is as follows:

- An own benchmarked dataset for suspicious outdoor activities in low light condition using infrared vision cameras may be created, focusing on some suspicious activities like fence climbing, fence wire cutting and unusual movements of armed personal.
- Further, new deep learning based models can be proposed for benchmarking of the dataset.
- Some other modalities may also be explored for accurate and efficient recognition of the activities.
- Deep learning models like Resnet-inception v2, inception v4, and more deeper networks of resnet with 200 and more layers may be implemented for better recognition accuracy.
- Different variations of deep learning model densenet (densenet-121, densenet-169 and densenet-201) may also be examined for performance improvement.