This chapter introduces first the background of video sequence analysis, which includes the need of video sequence analysis, applications, and difficulties involved therein. Thereafter, we describe one of the important applications of video sequence analysis that is human activity recognition which includes the systematic flow diagram of human activity recognition system, various steps used like segmentation, feature extraction and representation, and classification. The challenges involved in each steps are also discussed. At last, the significance of study and overview of the thesis is explained.

1.1 Video Sequence Analysis

A sequence of images displayed at a certain rate of frequency (frame per second) is termed as a video sequence. It consist information in the form of spatial changes with respect to time and if someone wants to extract the information from video sequence then the spatial change with respect to time must be perceived. For the video sequence analysis (VSA), the knowledge of core technologies of digital image processing is the fundamental requirement, which includes image enhancement, image segmentation, morphological operations, feature extraction and representation, image classification etc. The main aim of this analysis is to automatically detect and determine the spatio-temporal event in the video signal and further to interpret the nature of event. In this analysis, the identification of appropriate objects/regions in the scene (segmentation), and most descriptive characteristics of each object/region in the complete scene (feature
extraction) is extracted. In both the cases, the segmentation and feature extraction the spatial and temporal dimensions must be taken into consideration for effective representation of object/region. Feature extraction is useful for both coding and indexing and the most adequate coding parameters for each of the object/region is set. A similar process is used for segmentation, to identify the objects/regions, which permits particular and distinct coding and unrolls enhanced interaction possibilities. Similarly, the capability to define content in an object/region-based technique increases the plenty of the description.

Since last few decades, the VSA using intelligent techniques has emerged as a promising and interesting area of research in the field of computer vision and image processing due to the critical issues and numerous applications [1-3] of this analysis.

1.2 Critical Issues in Video Sequence Analysis

In the VSA, there are numerous factors which limit the performance of VSA system and these are the recording settings, illumination variations, camera motion, view point variations, complexity of the background, similarity between foreground and background object, high dimensionality and redundancy of the data etc [4]. All these factors provide an open challenge to the researchers/technocrats, to design and develop such an algorithm which has the capability to deal with these issues.

The environmental conditions play a very important role when the recording/acquisition of the video signal is done because the performance of vision based system is highly dependent on the weather conditions. The captured video signal in bad
environmental conditions leads to a poor quality of video signal. In the poor quality of video signal the object and background of the scene may not discernible and due to this the subsequent task (segmentation, feature extraction) associated in video sequence analysis system leads to worst performance [5]. Hence, proper illumination is needed to acquire good quality video signal, where object and background must be discernible.

The motion of camera creates blurred object in the scene and due to this the additional de-blurring algorithm is needed to de-blur the object [6]. Hence, to avoid this there must be a proper installation of camera.

The complexity of the background introduces the problem of extracting object from the scene. Due to cluttered background the object (foreground) and background may have more similarity [7] and due to this the accurate segmentation of object may not be possible. Hence, it is worthwhile that the recording of the scene must be performed based upon the application to avoid these issues.

1.3 Applications of Video Sequence Analysis

There are numerous applications of VSA and these are broadly categorised as:

- **Entertainment**: One of important application which is directly related to the daily life of human beings like television (TV), movies, high definition television (HDTV) transmission, video games, live streaming of sports analysis etc [8-10].
CHAPTER 1: Introduction to Video Sequence Analysis

- **Commercial:** VSA can be used for commercial purposes such as smart CCTV, advertisement of the product, in retail industry for tracking of shoppers inside the store etc [11].

- **Security and Surveillance:** Most widely used application for the security and safety purpose by military and police [12]. It can be monitoring of crowd behaviour at public functions, terrorist activities [13] at public places like airports, railways stations, bus stands etc., robbery detection, home intrusion system etc.

- **Human Computer Interaction:** Now a days the interaction of human with machine is increasing rapidly [14] due the advancement of VSA technologies. The traditional way of interaction of human with machine is through remote control, keyboard, mouse, joystick etc. but in the coming years these mode of interactions may become obsolete due to the invention of various recognition systems based on [15-16] body pose, hand gesture, facial expression, etc.

- **Motion Analysis:** There are a variety of systems, where VSA is used to detect and determine the motion of the object. The effective detection, tracking, and recognition of an object leads to several important applications like human activity recognition system [17] for detecting various kind of human abnormal and normal activities, object tracking, intruder detection and industrial monitoring.
As it is highlighted that VSA has number of applications in the various fields of science and technology but the main focus of this research is to design and develop a novel VSA algorithm for human activity recognition (HAR) system.

### 1.4 Human Activity Recognition System

A vision based human activity recognition system is capable of automatically detecting and determine the ongoing activity in the video sequence by extracting and interpreting the spatio-temporal change in the video sequence. As it is depicted in Figure 1, the HAR systems in general have some key processing steps like pre-processing, feature extraction and representation, and classification or recognition of an activity.

![Figure 1.1 Overview of HAR system](image)

In recent years, the area of vision based Human Activity Recognition (HAR) has grown dramatically, reflecting its importance in many high impact social applications including intelligent surveillance, web video search and retrieval, elderly care system for quality of life, content based video analysis, interaction between people, sports analysis, intelligent robotics, and prevention of terrorist activities [18-21]. The typical task of a HAR system is to detect and analyse human activity in a video sequence. The
reviews of previous work [22-23] reveals the challenges in vision based HAR systems. Various factors that make the task challenging are the variations in body postures, the rate of performance, lighting conditions, occlusion, view point and cluttered background. A good HAR system is capable of adapting to these variations and efficiently recognizes the human activity class. The important steps [24] involved in HAR systems are usually: a) Segmentation of foreground b) Efficient extraction and representation of feature vectors, and c) Classification or recognition. An effective and novel solution can be proposed at any step of the work individually, or collectively for all the steps. Due to the variation in human body taxonomy and environmental conditions, every step is full of challenges and therefore, one can only provide the best solution in terms of recognition accuracy and processing speed. The shape and motion feature based descriptors [1] are two widely used methods in HAR systems. Shape based descriptor is generally represented by the silhouette of the human body and silhouettes are the heart of the activity. Motion based descriptors are based on the motion of the body, and the region of interest can be extracted using optical flow and pixel wise oriented difference between the subsequent frames. The motion based descriptors are not efficient, especially when the object in the scene is moving with variable speed. The subsequent section highlights the details of issues related to HAR system and its functioning.

1.4.1 Major Challenges of Human Activity Recognition System

The major challenges of vision based HAR system is to deal with clutter background, motion of the camera, view point variations, illumination change, occlusion, intra-class
dissimilarity and inter-class similarity etc. The cluttered background creates a difficulty for selection of foreground object (segmentation) because of the disorderly arrangement of the objects in background and the object present in the background and foreground may have similar characteristics. For the segmentation of human silhouette, a prime step is to extract the human silhouette accurately for effective representation of human activity. Due to the sudden motion of the camera and view point change, the captured image is distorted, therefore it is vital that the position of camera should be fixed otherwise additionally process is required to detect the motion of the camera and view point change [25], which may be a complex task. The performance of vision based system is highly affected by the lighting conditions. Due to poor lighting condition the captured video signal has less variation of intensity of pixels and extraction of desired object is become a difficult task, hence it become an essential requirement to maintain proper illumination throughout the day but it is practically always not possible especially for the case of surveillance application. Therefore, again it becomes necessary to maintain proper illumination, where the HAR system is installed, but up to a certain level of variation of intensity of pixel can be adjusted by using proper enhancement techniques [26]. Due to the occlusion of object, the complete information about the object cannot be captured by single camera and the reason for occlusion in HAR system may be due to self-body part or alignment with other person. The effect of occlusion can be restricted in HAR system by incorporating multiple camera, which gives information at multiple views [27], or by employing a robust technique [28] for feature extraction but these solutions are having their own limitations. Having multiple cameras increases the cost and complexity of system while robust feature exaction
technique may be suitable only to track single person activity. Inter-class similarity and intra-class dissimilarity of different kinds of human activities plays a crucial role for the classification. To improve the classification accuracy of the HAR system a robust classifier which have the capability to deal with the high interclass similarity and intra-class dissimilarity [29].

1.5 Problem Statement

Video sequences of different human activities are given, which have clothing variation, zoom in, zoom out, illumination variation, high intra-class dissimilarity and inter-class similarity. Under these circumstances, we propose a solution of frameworks, detect and determine the human silhouette of the activity and based upon the human silhouette configuration, human activity is recognized and classified. A texture based entropy model provides the solution of accurate silhouette extraction under such variations. The problem of redundancy, losing geometrical and temporal information of feature representations are solved through key pose selection, cells and grid, computation of spatial distribution and the sum of directional pixels, and \( \mathcal{R} \)-transformed. The problem of classifier to deal with interclass similarity and intra-class dissimilarity, a robust multi-classifier is designed using the fundamentals of support vector machine.

1.6 Main Contribution of the Thesis

The main contribution of this thesis is to design and develop various novel approaches for the improving recognition accuracy of the HAR system. This thesis also gives the theoretical basis for the improvement of performance of HAR methods.
1.6.1 Theoretical Formulation

- The problem of segmentation, and spatial and temporal redundancy of the video signal have been identified.

- The issues involved in the recognition of human activity under the various lighting conditions have been featured.

- An issue related to the representation of 3-dimensional video signals has been identified and dealt with.

- Computation of the motion based temporal information of moving human is exhibited.

- The performance of the classifiers under various constraints of the activity performed have been observed.

- The less recognition accuracy of the HAR system under complex human activity has been detected.

1.6.2 Experimental Validation

- The issue of segmentation of object in a video sequence has been addressed by using a texture based entropy model.

- The spatial and temporal redundancy have been reduced by selecting key poses of human silhouette of an activity using highest energy concept.
• The representation of 3-dimensional video signal is done via 2-dimensional approach by forming average energy silhouette images (AESI).

• The loss of motion temporal information in AESI is boosted by computation of orientation of silhouette using $\Re$-transform.

• The effect of computation of spatial distribution of gradients at various levels are computed and validated using standard activity datasets.

• The robustness of the proposed algorithms are validated using standard datasets.

1.7 Motivations of Human Activity Recognition

The main motivation behind the study of HAR system is its huge applications in real world and the critical issues. Human activity recognition is a multidisciplinary area of research which associates with neural network, machine learning, intelligent computing, human computer interaction, as well as psychology and sociology. Thus, this field is drawing the attention of researchers for a variety of applications. Another important fact behind this work is to discover a novel framework for HAR system, which gives high recognition accuracy and in many literature it has been observed that multiple features [30-33] based HAR system gives an improved performance in comparison with single feature based system. Hence, in this work, it is also proposes some multiple features based approaches for the human activity recognition.
1.8 **Significance of the Study**

The core finding of this study leads to opening of a dam for many real life applications, which are entirely based on human activity recognition. Recently, it has been seen that the HAR system demands are increasing day by day due their potential real world applications like unconstrained video search, aerial video analysis, sport video analysis, health care system, gait analysis and biometric recognition, intelligent robot etc. Another important significance is the future research direction, which may ignite to the research community for further studies in this area with the help of current state-of-the-art.

1.9 **Thesis Overview**

In Chapter 2, the details of the earlier state-of-the-art methods, which includes their merits and demerits in terms of pre-processing, feature extraction and representation, and classification are described. It also gives the highlights the of research gap in the concerned area and based upon the research gap, the objectives of research is formulated and explained.

The silhouette and cells based bags-of-word model of activity recognition is explained in Chapter 3, which includes the detailed description of pre-processing (segmentation), feature extraction and representation, classifications, experimental setup and discussion of results.
In chapter 4, the activity recognition based upon the computation of spatial distribution of gradients, sum of direction pixels variations on average energy silhouette images, and human silhouette orientation based motion information computed using $\mathcal{R}$-transform are presented. The detailed analysis of the computation of spatial distribution of gradients (SDGs) on average energy silhouette images at different decomposition levels and further to propose an effective model of SDGs computation, which is experimentally verified and validated on standard human activity datasets.

Chapter 5 highlights the important conclusions drawn from the research, and also gives the details of future scope of work.