Chapter 1 : INTRODUCTION

This chapter presents our motivation behind the present work, an introduction to the problems discussed in this thesis, and objectives of this thesis. Finally the chapter concludes with a list of contributions of this thesis in field of video processing.

1.1. Motivation

Recent advances in digital imaging technology, computational speed, storage capacity, continuous increase in computational capabilities of hardware at low cost, and networking have made it possible to capture, manipulate, store, and transmit images and videos at interactive speed with equipment available at home or business. As a result, images and videos are now becoming an integral part of our day-to-day life, and are being used for entertainment, education, medicine, security, science and a number of other applications. Computer vision aims to provide description, understanding and interpretation of a scene by extracting the image features. There are a number of computer vision applications, which are used in a multitude of problems. For example, medical imaging, tumor detection, measurement of size and shape of internal organs etc. In forensics, it is desired that identity of a person would be ensured. In intelligent video surveillance, it is highly desirable that an intruder can be tracked and recognized. There are other important applications like industrial automation, radar imaging, remote sensing, robotics, etc. Now-a-days research activities are getting focused on development of intelligent video surveillance system with advances in computer vision methodology and processing capabilities. After terrorist attacks on 9/11 in USA, and that on 26/11 in Mumbai, India, a strong need for improvement in existing video surveillance capabilities was realized by the security establishments across the globe to prevent any such further terrorist attacks. Therefore, challenges and problems associated with development of intelligent video surveillance system needs to be

addressed. These challenges and problems motivated me to attempts three different problems namely – moving object (human, vehicle etc.) segmentation, background modeling and shadow suppression under highly dynamic background (such as bright sun, fog, heavy rain) are addressed and finally the human activities in multi-view camera are recognized.

1.2. Background

This thesis deals with three different problems for intelligent video surveillance system namely– moving object segmentation, background modeling and shadow suppression in complex wavelet domain under highly dynamic background, and finally the human activities recognition in multi-view camera. Intelligent video surveillance applications extract important features from video data, which provide a description, interpretation, or understanding of scene. The usual input to a surveillance system is a temporal image sequence with a corresponding amount of processed data.

In the last few decades, the wavelet based algorithms have been successfully used in different image and video processing tasks. The two milestone papers in late 1980's by Mallat [1] and Daubechies [2] have established baseline for applications of wavelet techniques. The wavelet transform is a powerful tool because wavelet representation is capable to illustrate both transient and stationary behavior of images with few wavelet coefficients. Discontinuities in an image often carry relevant information of image and therefore, they represent a critical part to be analyzed. However, a real valued wavelet transform suffers from three serious drawbacks [3-5]: (i) it is shift-sensitive in nature, (ii) it is having poor directional selectivity, and (iii) it provides no phase information. Due to these drawbacks real-valued wavelet transform is not suitable for video applications, because in video, the object may be present in translated and rotated form in different frames of the video and coefficients of real valued wavelet transform

corresponding to object region change abruptly across different frames due to its shift variant nature.

Complex wavelet transform [5-9] can reduce the above mentioned shortcomings of real-valued wavelet transform. Complex wavelet transform has not been explored to its potential. From the implementation point of view, complex wavelet transform can be implemented in different ways. Kingsbury [3] introduces Dual tree implementation of complex wavelet transform, which is not a complex wavelet transform in true sense. Lawton [8] and Lina [9] have proposed complex Daubechies wavelet transform, which is a natural extension of real Daubechies wavelet transform and is found useful in certain image and video applications. Daubechies Complex wavelet transform is most suitable to solve problems such as moving object segmentation, background modelling and shadow suppression.

Moving object segmentation is an important step for development of any intelligent video surveillance system [10]. Segmentation of moving object is a process of isolating the foreground object from background of a video sequence. Due to dynamic environmental conditions such as illumination variations, changing background, abrupt motion of object, and presence of shadow, moving object segmentation is a difficult and complicated task. As compared to the number of literature available on image segmentation [11], very limited literature is available on moving object segmentation. Commonly used techniques for moving object segmentation are background segmentation [12], statistical models [13], temporal differencing [14], and optical flow [15]. Most of the work on moving object segmentation have been done using spatial domain techniques, which suffer from problem of either inaccurate segmentation due to non-removal of noise or failure to detect new appearance automatically. Real valued wavelet transform based moving object segmentation methods [16, 17] have been

proposed to reduce the problems of spatial domain techniques. But as mentioned in preceding paragraph the real valued wavelet transform is not suitable for video applications, therefore successful exploitation of complex wavelet transform might reduce the different complicated tasks of moving object segmentation problem.

Dynamic background modeling and shadow suppression is a challenging problem. Background modeling is currently used to detect moving objects in video under critical situations like illumination changes, objects being introduced or removed from the scene. Many background modeling methods have been developed [18, 19] and these methods can be classified in the following categories: basic background modeling, statistical background modeling, fuzzy background modeling, and background clustering. The main objective of shadow detection and removal is to locate shadow regions, distinguish shadow from foreground object and then to remove it from scene. Shadow of objects create serious problems in different applications, due to misclassification of shadow points as object points. The real challenge is to distinguish object and its shadow in a video. Several dynamic background modeling and shadow suppression algorithms are available in literature for static background [20-21], whereas, works reported on background modeling and shadow suppression for moving object are limited. Out of these, some algorithms fail to give accurate results in various lighting and environmental conditions whereas others are not able to detect shadow accurately. Most of the approaches for background modeling and shadow suppression in moving object are in non-wavelet domain. Therefore, there is a strong need to develop algorithms in wavelet domain for background modeling and shadow suppression accurately. Accuracy and efficacy of these algorithms is also critical for activity recognition.

An activity is a sequence of movements generated during performance of a task. Recognition of activities provide important cues for human behavior analysis. Although a

large amount of work has been performed on activity recognition in the last few years yet still it is an open and challenging problem. The various issues and challenges involved in automatic human activity recognition from video sequences are as follows:

- The trajectory of activities from different viewing directions is different and some of the body parts (part of hand, lower part of leg, part of body, etc.) are occluded due to view changes.
- The other common issues include fixed or moving cameras, scenes having moving or clutter backgrounds, changes in light and view-point, variations in scale, starting and ending state, variations in appearance of individuals and cloths of human etc. These issues and situations make the human activity recognition a challenging task.
- Human activities are performed in real 3D environment and cameras only capture the 2D projection of the real scene. Therefore, visual analyses of activities carried out in the image plane are only a projection of the real activities. This projection of the activities depends on the viewpoint and do not contain full information about the performed activities.

Several human activity recognition methods were proposed in the past few years. Detailed surveys can be found in [22], where different methodologies of human activity recognition are discussed. Most of the work on activity recognition are view dependent and deal with recognition from one fixed view. Recognizing human activities from multiple views has been a challenging task for researchers around the globe and needs a lot of improvement. Inaccurate activity recognition of human objects leads to a failure in multiple human object activity recognition, therefore, multiple human object activity recognition is more complicated task than single human activity recognition.

1.3. Thesis Objectives

The objective of the proposed work in this thesis is to apply complex wavelet transform on images as well as on videos for different problems viz. moving object segmentation, background modeling and shadow suppression for different applications of intelligent video surveillance. Finally, in this thesis single and multiple human activities in multi-view camera are recognized.

The objectives of this thesis are as follows:

(i) Comprehensive literature review and comparative study of various classical as well as state-of-the art methods for moving object segmentation under varying illumination conditions. Further, design of new and efficient algorithms for moving object segmentation.

(ii) The design of a new and efficient algorithm for moving object segmentation using improved approximate median filter based method in complex wavelet domain.

(iii) To design new and efficient algorithm for background modeling and shadow suppression for moving object segmentation in complex wavelet transform, which makes the algorithm to work not only in case of normal video but also in case of maritime video with dynamic background and poor contrast.

(iv) To design new and efficient algorithms for multi-view human activity recognition for single person based on spatio-temporal template.

(v) To design new and efficient algorithm for multi-view human activity recognition for multiple person based on multiple features.

1.4. Outline of the Thesis

This thesis consists of seven chapters. Outline of the thesis is as follows:

Chapter 1 presents a brief introduction of the problems addressed in this thesis followed by the objectives of the thesis. Finally chapter concludes with a brief account on contributions of this thesis in field of image/video processing.

Chapter 2 discusses the theoretical background for video surveillance system. In this chapter, we have also given an overview of complex wavelet transform. Further, in this chapter a literature survey of prominent approaches for moving object segmentation and activity recognition are given.

Chapter 3 is organized in two-fold. In first fold, it presents a comprehensive literature review and comparative study of various classical as well as state-of-the art methods for moving object segmentation under varying illumination conditions. Further, a new method for moving object segmentation which is based on improved approximate median filter using Daubechies complex wavelet transform coefficients is developed. The proposed method and other state-of-the-art methods have been tested on different types of videos viz. normal video, noisy video, non- stationary background video and low quality with poor contrast video. Quantitative performance of the proposed method has been evaluated and compared with other state- of-the-art methods in terms of different performance metrics viz. average difference, structural content, normalized absolute error, misclassification penalty, percentage of correct classification, pixel classification based measure, relative foreground area measure and relative position based measure.

Chapter 4 presents a method based on complex wavelet transform for dynamic background modeling and shadow suppression of object in video. This chapter presents two methods for dynamic background modeling and shadow suppression in the complex

wavelet domain. First method handles the small movements of non-static objects such as tree branches and bushes blowing in the wind, waving trees, shadow regions that are projected by foreground objects and are detected as moving objects. Second method deals with highly dynamic background such as moving object in water surface, boat wakes, and weather issues (such as bright sun, fog, heavy rain), moving object in rain fall, and maritime object detection in night. Robustness of the proposed methods in comparison to other state-of-the-art methods has been tested for different types of videos. Finally, the quantitative performance values for different cases have been computed for the proposed method and other state-of-the-art methods and then compared in terms of different performance evaluation metrics.

Chapter 5 addressed the problem of multi-view human activity recognition. In this chapter, a multi-view human activity recognition system based on spatio-temporal template is proposed. Performance of this method has been analyzed in terms of confusion matrix and recognition accuracy.

Chapter 6 presents a method for multi-view activity recognition for multiple human. In this chapter a multi-view human activity recognition method is proposed by using multiple features i.e. (i) Contour based distance signal feature (ii) Uniform rotation local binary patterns. Further, the activities are classifying by using multiclass support vector machine (SVM) classifier. Performance of the proposed method has been analyzed in terms of confusion matrix and recognition accuracy.

Chapter 7 presents conclusions of the thesis and summarizes main findings of this thesis work. This chapter also proposes some possible future perspectives of the thesis.

1.5. Contributions to the Thesis

This section describes the important contributions of the thesis in the field of image/video processing which are as follows:

(i) Exploited different properties of complex wavelet transform, and demonstrated their suitability and application to solve different problems such as: moving object segmentation, background modeling, detection and removal of shadow from object in video.

(ii) A comprehensive literature review and comparative study of various classical as well as state-of-the art methods for moving object segmentation have been presented. Further, a new improved approximate median filter based method for segmentation of moving object in complex wavelet domain is proposed. The proposed algorithm is capable to segment the moving object for normal video, noisy video as well as for non-stationary background video.

(iii) Proposed two new methods for background modeling and shadow suppression for moving object segmentation using complex wavelet transform, which makes the algorithm to work not only in the case of normal or noisy video but also in the case of maritime video with dynamic background and poor contrast.

(iv) Proposed a new method for multi-view human activity recognition for single person. The proposed method is based on motion history images and object shape information.

(v) Exploited different features such as contour based distance signal feature and uniform rotation local binary patterns to solve multi-view human activity recognition for multiple person.