Chapter 2 : THEORETICAL BACKGROUND

This chapter discusses about theoretical background for video surveillance system. In this chapter, we have also given an overview of complex wavelet transform and its properties which provide basis for computer vision applications discussed in subsequent chapters of the thesis (e.g. moving object segmentation, background modelling & shadow suppression). Further, in this chapter a literature survey of prominent approaches for moving object segmentation and activity recognition are given. Section 2.1 presents the Introduction for video surveillance system. 2.2 presents the detailed discussion about complex wavelet transform. Section 2.3 presents the literature survey of moving object segmentation methods with their benefits and limitations. Section 2.3 presents the literature survey of activity recognition methods. Section 2.4 presents the detailed discussion about various performance measures used for qualitative and quantitative analysis.

2.1. Introduction

The goal of video surveillance is to detect and track certain moving objects in order to analyze their activities. Video surveillance is seeking the interest of the research community because of its many useful applications such as, crowd flux analysis, traffic monitoring, and access control to specific areas, human activity recognition, anomaly detection, and human computer interaction etc., [23-24]. The technological evolution of video-based surveillance systems started with analogue CCTV systems. The different surveillance systems are categorized under three generations. The first generation includes the analogue CCTV camera based surveillance systems. Second generation includes semi-automatic CCTV camera based systems assisted by computer vision algorithms. The third generation (modern generation) features automatic, real-time, distributed and multi-modal surveillance systems. The research of modern day is focused

for the development of efficient systems and the area is seeking attention of the researchers from various disciplines such as computer vision, behavioral sciences, signal processing, artificial intelligence and data mining.

Video surveillance has three main tasks (see Fig. 2.1), (a) moving object segmentation, (b) background modelling & shadow suppression for moving object, (c) object activity recognition. Various methods are found, in literature, for moving object segmentation, background modelling & shadow suppression for moving object, and object activity recognition. Most of the work on moving object segmentation, background modelling and shadow suppression 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. To avoid these shortcomings, the complex wavelet transform such as Daubechies complex wavelet transform seems to be a good candidate as Daubechies complex wavelet transform has no redundancy, has approximate shift invariance, allows perfect reconstruction and needs lesser computation as compared to other complex wavelet transforms. Daubechies complex wavelet transform can be implemented using linear phase filters and thus retains the shape of signal efficiently during the reconstruction process. Imaginary components of Daubechies complex wavelet coefficients contain edge information. The phase of wavelet coefficients carries skeleton of the image. Daubechies complex wavelet transform is useful for multiresolution analysis.

Activity recognition aims to recognizing actions of an actors in order to make inferences on high-level activities and goals by analyzing human behaviors as a series of observations.

Activity recognition is a difficult task because the shape of the different actors performing an activity can be different and the speed and style in which an activity is performed can

vary from person to person. Although the approaches to solve the activity recognition problem are few and far between, but the activity recognition methods available in literature can broadly be categorized into two groups: sensor based activity recognition and vision based activity recognition. In sensor based activity recognition methods some smart sensory device is used to capture various activity signals for activity recognition. Vision based activity recognition methods use the spatial or temporal structure of an activity in order to recognize it. A general framework for video surveillance system is given in Fig. 2.1

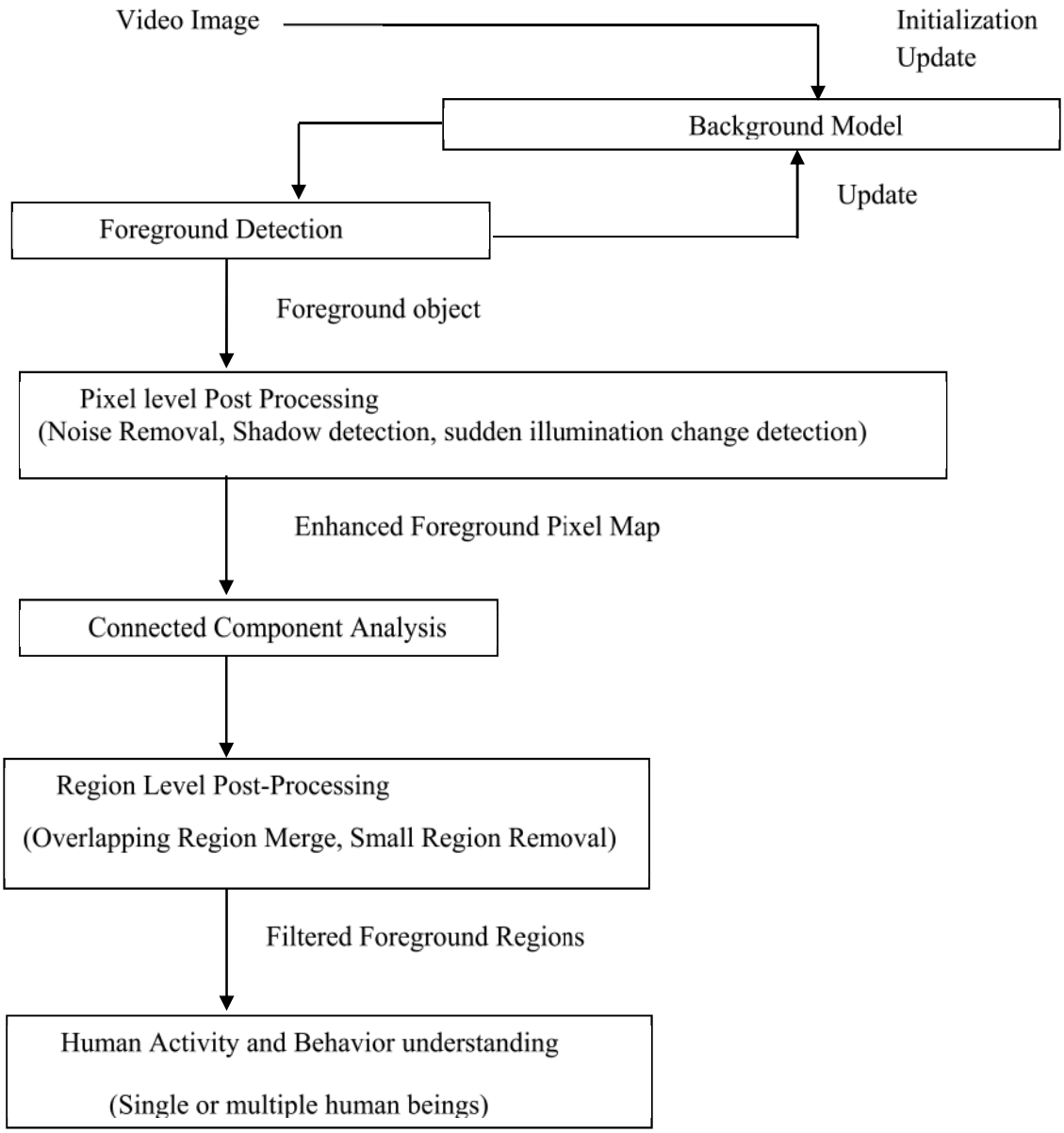

Figure 2.1: Block diagram for video surveillance system

2.2. Daubechies Complex Wavelet Transform

In this section, we have discussed Daubechies complex wavelet transform and its properties which provide basis for computer vision applications discussed in subsequent chapters of the thesis (e.g. moving object segmentation, background modelling & shadow suppression). In a video, an object may be present in translated as well as in rotated form among different frames. Most of the transform coefficients, including Discrete Fourier Transform (DFT) vary by translation and rotation of the object. By using Daubechies complex wavelet transform, wavelet coefficients corresponding to object region remain approximately invariant for translation and rotation of object.

2.2.1. Construction of Daubechies Complex Wavelet

The basic equation of Multiresolution theory is the scaling equation [4]

$$\phi(u)=2\sum_i a_i\phi(2u-i) \tag{2.1}$$

where a_i's are coefficients, and $\phi(u)$ is the scaling function. The a_i's can be real as well as complex valued. Daubechies's wavelet bases $\{\psi_{j,k}(t)\}$ in one-dimension is defined using the above mentioned scaling function $\phi(u)$ and multi-resolution analysis of $L_2(\Re)$ [4]. The generating wavelet $\psi(t)$ is defined as:

$$\psi(t)=2\Sigma(-1)^n\,\overline{a_{1-n}}\,\phi(2t-n) \tag{2.2}$$

where $\psi(t)$ share same compact support [-L, L+1].

Any function f (t) can be decomposed into complex scaling function and mother wavelet as:

$$f(t)=\sum_k C_k^{j_o}\phi_{j_o,k}(t)+\sum_{j=j_o}^{j_{\max}-1} d_k^j\psi_{j,k}(t) \tag{2.3}$$

where, j_o is a given low resolution level, $\{C_k^{j_o}\}$ is called approximation coefficient and $\{d_k^j\}$ is known as detail coefficient.

2.2.2. Advantages of Daubechies Complex Wavelet Transform

Daubechies complex wavelet transform can be useful for several image and video processing applications as it owns following advantages:

(i) It is approximate shift invariant [6]. A transform is shift-sensitive if an input signal shift causes an unpredictable change in transform coefficients. In Discrete Wavelet Transform (DWT) the shift-sensitivity arises from down samplers in the implementation. Daubechies complex wavelet transform has reduced shift sensitivity. Shift invariance property [7] also affect the informational content at multilevel. In DWT as one moves toward higher level, the nature of wavelet coefficients at different sub-bands changes unpredictably while in the case of the Daubechies complex wavelet transform, the nature of wavelet coefficients at different sub-bands remains unchanged.

(ii) It has perfect reconstruction property.

(iii) Daubechies complex wavelet transform has no redundancy, as it uses a single filter bank whereas the dual-tree complex wavelet transform has a redundancy of $2^m:1$ for m-dimensional signal [3] because it uses two filter banks, one for real part and the other for imaginary part.

(iv) Computational requirement in Daubechies complex wavelet transform (although it involves calculation using complex numbers) is of the order of that in DWT, that is, the computational complexity is O (number of data points) whereas other complex wavelet transform, say DT-CWT has $2m$

times more computational requirement as compared to that in DWT for *m*-dimensional signal.

(v) It provides true phase information.

(vi) Use of Daubechies complex wavelet transform results in better edge detection as compared to use of real valued wavelet transform.

2.2.3. Properties of Daubechies Complex Wavelet Transform

In this section, we have shown different important properties of Daubechies complex wavelet transform which are useful in different applications discussed in this thesis. Properties of Daubechies complex wavelet transform are described as follows:

(i) Symmetricity and linear phase property

The symmetry property of filter bank makes it easy to handle the boundary problem for finite length signals. The linear phase response of the filter precludes the non-linear phase distortion and retains the shape of the signal which is very useful in imaging applications.

(ii) Edge detection property

Let $\phi(t) = k(t) + i\,l(t)$ be a scaling function. Let $\hat{l}(\omega)$ and $\hat{k}(\omega)$ are Fourier transforms of $l(t)$ and $k(t)$. Clonda *et al.* [5] considered that ratio $\alpha(\omega)$ for the scaling function as follows:

$$\alpha(\omega) = -\frac{\hat{l}(\omega)}{\hat{k}(\omega)} \qquad (2.4)$$

Clonda *et al.* [5] observed that $\alpha(\omega)$ is strictly real-valued and behaves as ω^2 for $|\omega| < \pi$. This observation relates the imaginary and real components of scaling function as $l(t)$ accurately approximates another derivative of $k(t)$, up to some constant factor.

As a result, it can be derived from Eq. (2.4) that $l(t) \approx \alpha \Delta^2 k(t)$, where Δ represents second order derivative. This gives multiscale projection as,

$$\begin{aligned} \left\langle f(t), \phi_{j,k}(t) \right\rangle &= \left\langle f(t), k_{j,k}(t) \right\rangle + i \left\langle f(t), l_{j,k}(t) \right\rangle \\ &\approx \left\langle f(t), k_{j,k}(t) \right\rangle + i\alpha \left\langle \Delta f(t), k_{j,k}(t) \right\rangle \end{aligned} \tag{2.5}$$

From Eq. (2.5), it can be concluded that the real component of complex scaling function carries averaging information and the imaginary component carries edge information.

Daubechies complex wavelet transform acts as local edge detector because imaginary components of complex wavelet coefficients represent strong edges. This helps in preserving the edges and implementation of edge sensitive computer vision applications such as moving object segmentation, background modelling and shadow suppression etc.

(iii) Availability of phase information

The phase of complex wavelet coefficients represents the skeleton of the signal. The discrete wavelet transform does not provide any phase information. Phase and modulus of complex wavelet coefficients collaborate in a non-trivial way to describe the data [5]. The phase encodes most of the coherent structure of the image while the modulus encodes mostly the strength of local information that could be corrupted with the noise [5]. The complete understanding of informational content of the phase of wavelet coefficients is still an open problem.

(iv) Approximate Rotation invariance property

Daubechies complex wavelet transform is approximate rotation invariant. A transform is said to be rotation invariant, if rotation in input signal causes same rotation in transform coefficients.

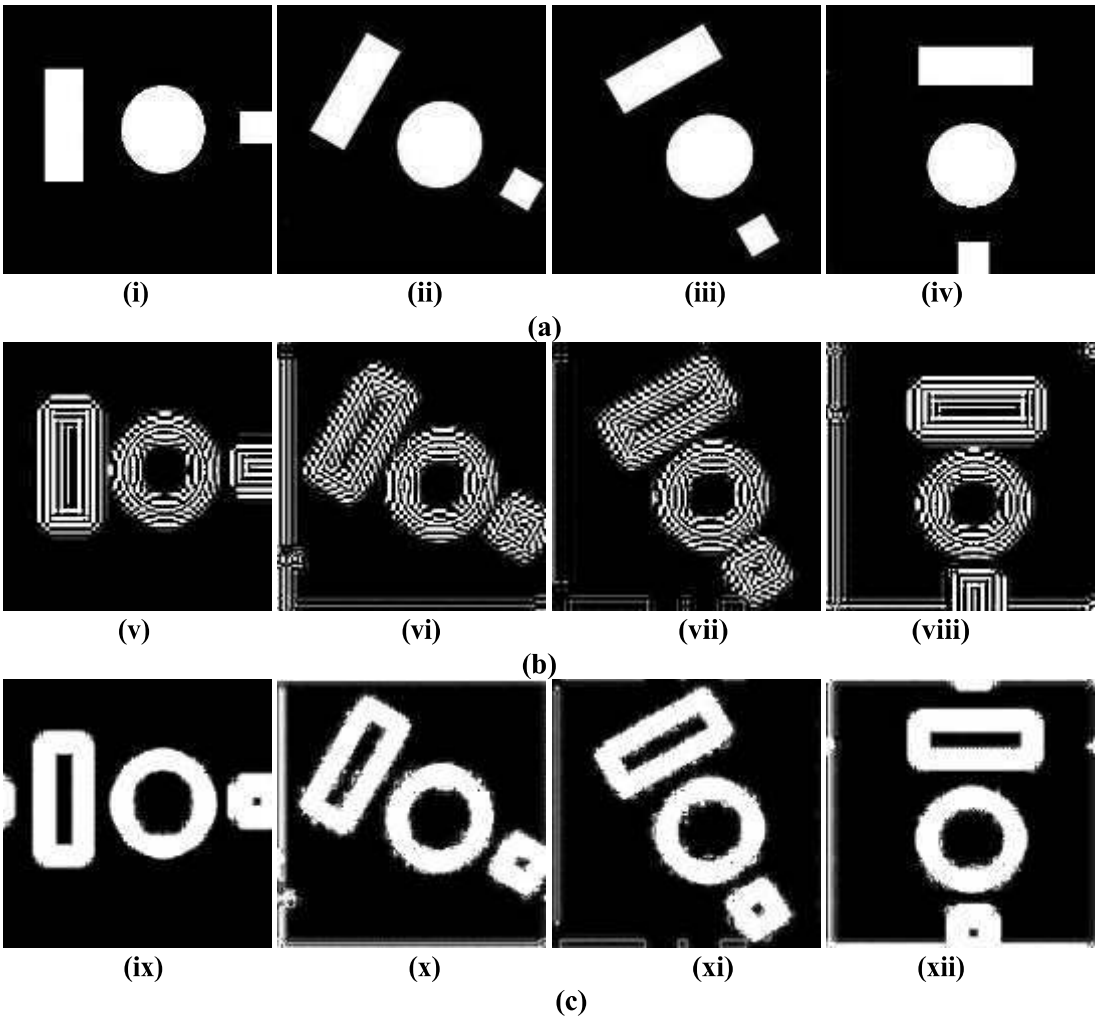

Figure 2.2: Rotation invariance property of Daubechies complex wavelet transform: (a) [(i)- original image, (ii)-rotation performed with 30 degree angle, (iii)-rotation performed with 60 degree angle, (iv)-rotation performed with 90 degree angle], Image reconstructed from 2 levels of wavelet coefficients using (b) discrete wavelet transform [(v)-(viii)], and (c) Daubechies complex wavelet transform [ix-xii].

Fig. 2.2 shows the approximate rotation invariant nature of Daubechies complex wavelet transform. Fig. 2.2 (i) -(iv) shows original image, image rotated by 30 degree, 60 degree and 90 degree angles respectively. Reconstructed images by discrete wavelet transform and Daubechies complex wavelet transform at 2 levels are shown in Fig. 2.2 (v) - 2.2 (viii) and 2.2 (ix) – 2.2 (xii) respectively. In Fig. 2.2, there are two types of object structures: rectangular edge and circular edge. As both types of object structure rotates through space with some angle, the reconstruction using real valued discrete wavelet transform changes erratically, whereas Daubechies complex wavelet transform

reconstructs all local shifts and orientations without significant variations. This indicates that Daubechies complex wavelet transform is approximate rotation-invariant.

2.3. Literature Survey of Moving Object Segmentation

Moving object segmentation is a crucial step in video analysis and is associated with a number of computer vision applications. Segmentation of moving object is a process of isolation of foreground object from background in a video sequence. Importance of the moving object segmentation problem is evident from the fact that it is essential building block for video surveillance, object detection and recognition, indoor/outdoor object classification and many other applications [23-24]. To design the moving object segmentation algorithm for video surveillance systems, several major challenges have to be concerned. Toyama *et al.* [25] have identified the following challenges in moving object segmentation such as (i) lighting changes, shadows and reflections (ii) dynamic backgrounds such as waterfalls or waving trees (iii) Motionless foreground (iv) small movements of non-static objects such as tree branches and bushes blowing in the wind (v) noise image, due to a poor quality image source (vi) movements of objects in the background that leave parts of it different from the background model (ghost regions in the image) (vii) multiple objects moving in the scene both for long and short periods (viii) shadow regions that are projected by foreground objects and are detected as moving objects. Out of all these issues, changing illumination conditions remain a major problem for moving object segmentation in real-life problems. To take into account these problems, many background modeling methods have been developed [18, 19] and these methods can be classified in many categories such as basic background modeling, statistical background modeling, background clustering, and wavelet based background modeling.

Basic background modelling technique is based on the usage of average pixels. Kushwaha *et al.* [26] proposed a segmentation method which is based on construction of statistical background model. In this method, variance and covariance of pixels are computed to construct the model for scene background which is adaptive to the dynamically changing background. Mcfarlane and Schofield [27] have proposed approximation median filter method for segmentation of multiple video object. In this method, at first the frame differences are obtained from background frame and next successive frames, in second and last step temporal updating of the background frame is performed to adapt the changes in background in lighting conditions. This method has also been used in background modelling for urban traffic monitoring [44] but it suffers from the problem of unwarranted ghost objects and shadows. Mei *et al.* [28] proposed automatic segmentation method for moving objects based on the spatial-temporal information of video. In this method, the author utilizes the spatial-temporal information. Spatial segmentation is applied to divide each image into connected areas to find precise object boundaries of moving objects. The limitation of this method is that the boundaries of the extracted objects are not always accurate enough to locate them in different scenes.

Statistical background modelling using a single Gaussian or mixture of Gaussians. Statistical variables are used to classify the pixels as foreground or background. Stauffer *et al*. [29] have proposed a tracking method; in this method motion segmentation has been done using mixture of Gaussians and use on-line approximation to update the model. Ellis *et al*. [30] have proposed online segmentation of moving objects in video using online learning. In this approach, motion segmentation is done using semi-supervised appearance learning task wherein supervising labels are autonomously generated by a motion segmentation algorithm but the computational complexity of this algorithm is very low. Wren *et al.* [31] have proposed Running Gaussian Average model for moving object

segmentation. The model is based on Gaussian probability density function (pdf). In this method, a running average and standard deviation are maintained for each color channel. Zivkovic [32], proposed a moving object segmentation technique which is combination of temporal and spatial features. This approach automatically adapts the number of Gaussians being used to model a given pixel.

In background clustering, incoming pixels are matched against the corresponding cluster group and classified according to whether the matching cluster is considered part of the background. Kim *et al.* [33] proposed motion segmentation method based on codebook. A codebook method is formed to represent significant states in the background using quantization and clustering [33]. Here, RGB color space is used for codebook formation. The codebook based technique suffers from the problem of unwarranted ghost objects and shadows. The method is adaptive to only the small and gradual changes in the background and in case of sudden changes it distorts. Liu *et al.* [34] has proposed inter-frame difference method for video segmentation which is based on cumulated difference, object motion and adaptive thresholding. Due to lack of spatial information of objects, these algorithms suffer from unwarranted ghost objects, shadows, changing background, clutter, occlusion, and varying lighting conditions [25].

Wavelet based background modeling is defined in the temporal domain, utilizing the coefficients of Discrete Wavelet Transforms (DWT). Huang *et al.* [16, 35] proposed an algorithm for moving object segmentation using single change detection and double change detection applied in wavelet domain. Baradarani [36, 37] refined the work of Huang *et al.* [16, 35] using dual tree complex filter bank in wavelet domain. However, it has several drawbacks. First, the value of the frame difference depends on the speed of the object, so the quality of the segmentation cannot be maintained consistently if the speed of the object changes significantly in the sequence. It has a relatively poor result

for video sequences like the surveillance type of video. To overcome these problems, Im *et al.* [38] used the background registration technique. Kushwaha *et al.* [39] also refine the work of Baradarani and Huang using Daubechies complex wavelet domain. This method may reduce noise disturbance and speed change, but the segmenting coherence is still not ensured [40].

2.4. Literature Survey of Activity Recognition Methods

Activity recognition is a popular area of research in the field of computer vision. It is the basis of many applications such as security, surveillance, clinical applications, biomechanical applications, human robot interaction, entertainment, education, training, digital libraries, video or image annotations, video conferencing and model based coding. Recognition of activities provides important cues for human behavior analysis techniques. 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 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.

To overcome these problems the idea of using information through the cameras placed at different views has been used [41-43]. Use of information obtained from multiple cameras at multiple views provides efficient analysis of actual human activities. Several human activity recognition methods have been proposed in the past few years. Detailed surveys can be found in [44], where different methodologies of human activity recognition are discussed. The standard approach for human activity recognition is to extract a set of features from each image sequence frame and use these features to train classifiers and to perform the recognition. Therefore, it is important to consider the appropriateness and robustness of features of activity recognition in varying environments. There are many sources of variability that can affect human activity recognition such as variation in speed, viewpoint, size and shape of performer, phase change of activity, scaling of persons, and so on. Among various features the motion and shape features of human body parts play the most significant roles for activity recognition.

Motion feature based activity recognition has been performed by several researchers, such as reported in papers [45-48]. However, most motion-based techniques are not robust in capturing velocity when the motions of the actions are similar to the same body parts. Bobick *et al.* [45], used motion templates for recognizing the activities in a specific environment of aerobic exercise. They used Motion Energy Image (MEI) for obtaining segmented foreground and Motion History Image (MHI) for obtaining motion information in a view-specific environment. It does not give good activity recognition accuracy in outdoor environment. A more refined application of this algorithm was proposed by J. Ben-Arie *et al.* [46] which used a set of the velocity vector to represent

each activity and store the representation of a set of multidimensional hash tables and use multidimensional indexing to recognize activities but its problem to distinguish the walking and running activities. To deal with the issues mentioned in [45-46], Sharma *et al.* [47] have proposed a human activity recognition method which used motion history images and object shape information for different human activities in a video. Hamid *et al.* [48] extract spatio-temporal features such as the relative distance between two hands and their velocities and use dynamic Bayesian networks to recognize human activities. The motion-based features can easily be used to discriminate between walking and sitting down, but it fails to discriminate between walking and slow running or jogging.

Shape-based activity recognition has been performed by several researchers, such as [49-51]. Cohen *et al.* [49] proposed a view-independent 3D shape feature for classifying the human posture using SVM classifier. Carlsson *et al.* [50] also proposed a method for activity using shape feature. In this method, the shape was represented by key-frames and edge data for each activity. Then, a shape matching algorithm was applied to localize key frames of new video sequences to recognize the different activities. Veeraraghavan *et al.* [51] used shape features and procrustes distance to recognize the human activity. One of the problems of this method is that the articulated, non-rigid nature of human actions can make the registration of the point sets pretty challenging.

Motion-based and shape-based features have their own limitations. The motion-based features can depict the approximate moving direction of the body, but most of the motion based features are not robust in capturing velocity. On the other hand, shape-based features can represent pose information of the body, but without motion information its capability of describing the human activity is limited. To concern these issues, many researchers proposed combined features of motion and shape.

Combined features of motion and shape based activity recognition have been reported by several researchers, such as [52, 53]. Ahmad *et al.* [52] proposed a human action recognition system using combined motion and shape flow feature which characterize motion and shape properties for recognition of human actions in daily life. In this approach, author does not recognize all the action in front-normal view. Althloothi *et al.* [53] proposed an approach for activity recognition using multi-features and multiple kernel learning. In this approach, two sets of features are used such as shape and kinematic structure for human activity recognition using a sequence of RGB-D images. The major disadvantage of this method is that it is view dependent and deals with activity recognition from one fixed view. Table 2.1 presents the summary of various human activity methods.

2.5. Performance Measures

In this section, the brief descriptions of parameters used to evaluate the performance measures of moving object segmentation, background modelling & shadow suppression, and activity recognition are discussed as follows:

- **Relative Foreground Area Measure (RFAM) [84]**

RFAM gives accurate measurements of the object properties such as area and shape, more specifically the area of the detected and expected foreground. It is calculated between ground-truth frame and segmented frame [84]. The value of RFAM is in the range [0, 1] and if it is 1 then it indicates that the chosen method perfectly segment the moving object. The RFAM is calculated as follows:

$$RFAM = 1 - \frac{\left| Area(I_{GTV}) - Area(I_{SEGM}) \right|}{Area(I_{SEGM})} \tag{2.6}$$

where $Area(I_{GTV})$ and $Area(I_{SEGM})$ are area in objects in ground –truth frame and resulting segmented frame, respectively.

Table 2.1: Summary of various human activity methods

Year	First author	Dim	Feature/Representation	Classifier/Matching
2005	K. Huang [54]	3D	3D shape context, tracking	Hidden Markov model
2006	Ahmad [55]	2D	Optical flow, PCA, Human body shape	Hidden Markov model
2006	Canton-Ferrer [56]	3D	3D Motion descriptors and invariant moments	Bayesian classier
2006	Pierobon [57]	3D	Cylindrical shape descriptor	DTW, template matching
2006	Weinland [58]	3D	Motion history Volumes (MHV), FFT	LDA, Mahalanobis distance
2007	Lv [59]	2D	Shape context, graph model: Action Net	Viterbi algorithm
2007	Weinland [60]	2D	Exemplars of silhouettes projections	Hidden Markov model, 3D learning
2008	Cherla [61]	2D	Width prole, Spatio-temporal features	DTW, average template matching
2008	Farhadi [62]	2D	Histogram of silhouette and optic flow	A transferable activity model
2008	Junejo [63]	2D	Bag of local self-similarity matrices	Support vector machines
2008	Liu [64]	2D	Local spatio-temporal volumes, spin-images	Fiedler Embedding
2008	Liu [65]	2D	Bag of local Cuboid features	Support vector machines
2008	Souvenir [66]	2D	R transform surfaces, manifold learning	2D diffusion distance metric
2008	Tran [67]	2D	Motion context	Nearest neighbor and rejection
2008	Turaga [68]	3D	MHV, Stiefel and Grassmann manifolds	Procrustes distance metric
2008	Vitaladevuni [69]	2D	Motion history images, ballistic dynamics	Bayesian model
2008	Yan [70]	3D	4D action feature model	Maximum likelihood function
2009	Gkalelis [71]	2D	Multi-view posture masks, DFT, FVT	LDA, Mahalanobis distance
2009	Kilner [72]	3D	Shape similarity	Markov model
2009	Reddy [73]	2D	Feature-tree of Cuboids	Local voting strategy
2009	Veeraraghavan [74]	3D	Circular FFT features	DTW, Bayesian model
2010	Huang [75]	3D	Shape histogram, shape-flow descriptor	Similarity matrix
2010	Iosidis [76]	2D	Multi-view binary masks, FVT	LDA, Euclidean/Mahalanobis dist.
2010	Weinland [77]	2D	3D Histogram of Oriented Gradients	Hierarchical classification
2011	Haq [78]	2D	Dynamic scene geometry	Multi-body fundamental matrix
2011	Holte [79]	3D	3D optical flow, Harmonic motion context	Normalized correlation
2011	Junejo [80]	2D	Bag of temporal self-similarities	DTW, Support vector machines
2011	Liu [81]	2D	Bag of Cuboids features	Knowledge transfer, graph matching
2011	Pehlivan [82]	3D	Circular body layer features	Nearest neighbor
2011	Song [83]	3D	3D body pose and HOG hand features	Hidden conditional random fields

- **Misclassification Penalty (MP) [84]**

The MP estimated in the segmentation results which are farther from the actual object boundary (ground-truth image) are penalized more than the misclassified pixels which are closer to the actual object boundary [84]. The MP value lies in the range [0, 1]. Zero value of MP means perfect segment the moving object and performance of segmentation methods degrades as the value of MP becomes higher. The MP is defined as follows:

$$MP = \frac{\Sigma_{(i,j)} X(i,j).\, Chem_{GTV}(i,j)}{\Sigma_{(i,j)}\, Chem_{GTV}(i,j)} \tag{2.7}$$

where $Chem_{GTV}$ denotes the Chamfer distance transform of the boundary of ground-truth object. Indicator X can be computed as

$$X(i,j) = \begin{Bmatrix} 1, & if & I_{GTV}(i,j) \neq I_{SEGM}(i,j) \\ 0 & if & I_{GTV}(i,j) = I_{SEGM}(i,j) \end{Bmatrix} \tag{2.8}$$

where $I_{GTV}(i,j)$ and $I_{SEGM}(i,j)$ are ground-truth frame and segmented frame respectively with dimension $(i \times j)$.

- **Pixel Classification Based Measure (PCM) [84]**

The PCM reflects the percentage of background pixels misclassified as foreground pixels and conversely foregrounds pixels misclassified as background pixels [84]. The values of PCM lies in the range [0, 1] with its value 1 indicating perfect moving object segmentation. The PCM is defined as follows:

$$PCM = 1 - \frac{Cardi(B_{GTV} \cap F_{SEGM}) + Cardi(F_{GTV} \cap B_{SEGM})}{Cardi(B_{GTV}) + Cardi(F_{GTV})} \tag{2.9}$$

where B_{GTV} and F_{GTV} denote the background and foreground of the ground-truth frame, whereas B_{SEGM} and F_{SEGM} denote the background and foreground pixels of the achieved

segmented frame '$\cap$' is the logical AND operation. $Cardi(.)$ is the cardinality operator.

- **Relative Position Based Measure (RPM) [84]**

RPM is defined as the centroid shift between ground-truth and segmented object mask. It is normalized by parameter of the ground-truth object [84]. The value of RPM is 1 for perfect moving object segmentation. The RPM is calculated as:

$$RPM = 1 - \frac{\left\| Cent_{GTV} - Cent_{SEGM} \right\|}{2\sqrt{\pi}.\sqrt{Area_{GTV}}} \tag{2.10}$$

where $Cent_{GTV}$ and $Cent_{SEGM}$ are centroid of objects in ground-truth frame and achieved segmented frame respectively. $Area_{GTV}$ is the area of object in ground-truth frame. $\|.\|$ is the Euclidean distance.

- **Peak Signal-to-Noise Ratio (PSNR) [87]**

The mean square error (MSE) and the peak signal-to-noise ratio (PSNR) are the two error metrics used to compare image compression quality [87]. Higher value of PSNR means good segmentation (i.e. noise is minimum) while low value of PSNR indicates poor segmentation (i.e. noise is maximum).

$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right) \tag{2.11}$$

- **Normalized Absolute Error (NAE) [86]**

NAE is calculated between ground truth frame and segmented frame [86]. Lower values of NAE indicate perfect moving object segmentation while high value of NAE indicates poor object segmentation.

$$NAE = \frac{\sum_{x=1}^{\alpha}\sum_{y=1}^{\beta}\left|F_{GTV}(x,y) - F_{SEGM}(x,y)\right|}{\sum_{x=1}^{\alpha}\sum_{y=1}^{\beta}F_{GTV}(x,y)} \tag{2.12}$$

Where F_{GTV} and F_{SEGM} are ground-truth frame and segmented frame respectively with dimension $(\alpha \times \beta)$.

- **Normalized Cross Correlation (NCC) [85]**

NCC can be used as a measure for calculating the degree of similarity between two images [85]. NCC value lies in the range [0, 1]. Higher value of NCC indicates better segmentation while lower value of the same indicates poor segmentation.

$$NCC = \frac{\sum_{J=1}^{M}\sum_{K=1}^{N} F(J,K)\hat{F}(J,K)}{\sum_{J=1}^{M}\sum_{K=1}^{N}[F(J,K)]^2} \tag{2.13}$$

where $F(J,K)$ is ground truth frame and $\hat{F}(J,K)$ is achieved segmented frame.

- **Shadow Detection Rate [88]:**

Shadow detection rate is defined as

$$\varkappa = \frac{TP_S}{TP_S + FN_S} \tag{2.14}$$

where, subscript S stands for shadow. The TP and FN are number of true positives and number of false negatives.

- **Shadow Discrimination Rate [88]:**

Shadow discrimination rate is defined as

$$\zeta = \frac{\overline{TP_F}}{TP_F + FN_F} \tag{2.15}$$

where, subscript F stands for foreground. The TP and FN are number of true positives and number of false negatives. The $\overline{TP_F}$ is the number of ground truth points of the foreground objects minus the number of points detected as shadows, but belonging to foreground objects.

- **Recall & Precision [89]:**

The precision and recall parameters are calculated as follows

$$\text{Recall} = \frac{TPR}{TPR + FNR} \tag{2.16}$$

$$\text{Precision} = \frac{TPR}{TPR + FPR} \tag{2.17}$$

Here TPR represents true positives rate, FNR represents false negatives rate and FPR represents false positives rate.

- **F-Measure [89]:**

The F-Measure is calculated as follows [32]:

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.18}$$

- **Correct Recognition Rate (CRR)**

CRR for activity recognition is defined as follows:

$$CRR = \frac{N_c}{N_a} \times 100 \,(\text{in percentage}) \tag{2.19}$$

where N_c is the total number of correct recognition sequences while N_a is the number of total activity sequences.

2.6. Conclusions

This chapter presented the theoretical background for video surveillance systems. In this chapter, an overview of complex wavelet transform and its properties was given which provided basis for computer vision applications discussed in subsequent chapters of the thesis (e.g. moving object segmentation, background modelling and shadow suppression). Further, in this chapter a literature survey of prominent approaches for moving object segmentation and activity recognition was given.