CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The computerized video classification has become more important due to the vast availability of digital video contents. A large amount of video data is accessible in today's world, both on the internet and television. This makes it very difficult to find a single video of interest from such voluminous content. Video classification was used to search the choice of a video inside a specific video category. Video collections can be categorized into news, sports, commercial, cartoon and so on. Among the various video collections, sports video classification is a vital application area due to their huge commercial appeal. To assess the present status of this research work, detailed literatures were reviewed and that are presented in this chapter.

2.2 REVIEWS ON SPORTS VIDEO CLASSIFICATION

Transform coefficients based video classification system was designed by Girgensohn and Foote (1999), composed of three sequential modules- feature extraction, feature reduction and classification. The statistical models of reduced Discrete Cosine Transform (DCT) or Hadamard transform coefficients were extracted from the video frames. The high dimensional feature space of the transformed coefficients may increase the complexity of the classification system. To avoid this, Principal Component Analysis (PCA) algorithm was applied. Finally, the reduced set

of features was fed into Gaussian Markov Model (GMM) classifier for video classification.

An unsupervised cluster analysis based automated video abstraction approach was presented by Hanjalic and Zhang (1999). The underlying principle of the video abstraction method was to remove the visual content redundancy among video frames. Initially, the entire video material was grouped into clusters, each containing frames of similar visual content. The clustering process would group together all frames from all shots belonging to each of the content components and resulting in only two clusters, one for each of the components. Then, by representing each cluster with its most representative frame, a set of key frames was obtained which summarizes the given sequence. To obtain a summarizing preview sequence, it was sufficient to take the shots to which the extracted key frames belong as key segments and to concatenate them together.

Content based approach was employed for video classification scheme by Truong and Dorai (2000). It consisted of two stages- feature extraction and classification. In the feature extraction stage, three types of features: editing, motion and color features were extracted. Editing was a useful feature in video characterization. The percentage of each type of transition in video clips was used to extract editing features. The camera motion influences the narration of scene content. Hue Saturation Value (HSV) color model was adopted for color feature extraction. In the final stage, decision tree was built to classify the unknown videos.

Video classification based on dynamic content of short sequences was implemented by Roach et al. (2001). The dynamic contents were extracted from the

background camera motion and foreground object motion. The extracted motion undergoes into feature extraction by employing DCT and GMM classifier was used for classification. An approach for event detection and summarization was described by Li and Sezan (2001) for three sports video such as American football, baseball, and sumo wrestling. Generally, the important events in a class of sports were modelled by 'plays'. In order to detect play sequences, deterministic and probabilistic approaches were used. Then, the detected plays were concatenated in order to generate a compact, time compressed summary of the original video. The concatenated videos were used for best higher level summarization and analysis than the original sports video.

A semantic characterization approach for sports video annotation was illustrated by Assfalg et al. (2001). Videos were automatically annotated according to its visual content at different layers of semantic significance. Then, each shot was decomposed into its visual and graphic content elements, including foreground and background, objects, text captions, etc. In order to capture semantic content at a higher level of significance, several different low-level visual primitives were combined together by domain specific rules.

A framework for scene detection and structure analysis in sports video was implemented by Zhong and Chang (2001). Various features such as domain specific knowledge, supervised learning techniques, automatically segmented objects and features from both compressed and uncompressed domains were combined for scene detection and structure analysis process. The experiments were carried out for tennis and baseball sports video. Motion pattern descriptors based video classification system was designed by Ma and Zhang (2002). Motion texture was an efficient

representation of motion patterns, which were derived from the motion field in either optical flow or Motion Vector Field (MVF) in MPEG video between video frames. Motion texture extracted textures from MVF as motion pattern descriptor. In order to classify sports video into one of the predefined sports category kernel SVM classifier was adopted.

Acoustic and visual descriptor based automated non-editing broadcasted video classification was demonstrated by Roach (2002). Low level holistic motion measure was extracted as visual feature. Discrete Fourier transform based Mel Frequency Cepstral Coefficients (MFCC) was considered as acoustic features. GMM classifier was adopted for video classification by exploiting extracted features. The performance of the system was evaluated by acoustic and visual feature alone and a linear fusion of both features. An approach for automated sports video classification was presented by Messer et al. (2002). It relies upon the concept of "cues". On account of automated classification of sports video, low level-features were computed from the given video.

Multimedia features based baseball scene classification was illustrated by Hua et al. (2002). The classification of video scenes was important for video summarization, archiving and indexing. The integration of multimedia features, including image, audio and speech cues were considered. In order to select the multimedia features from temporal contexts, the maximum entropy scheme was used. TV broadcast baseball videos were used to perform experimental evaluation.

An automated video classification approach using decision tree classifier was illustrated by Yuan et al. (2002). At first, key frames were selected from the given

video using largest color histogram difference with the previous frame in the shot. Then, color and motion features such as the average color luminance of the whole video in RGB (Red, Green and Blue) color space, average standard deviation, camera motion, object motion, optical flow, and dynamics were extracted from the selected key frames. Finally, the extracted features were trained by the decision tree classifier for sports video classification.

Video shot classification was described by Yu et al. (2002) for tennis based on domain independent features and domain dependent features. By fusing these two features along with the domain knowledge, tennis video shots were classified into five classes. A generative model approach for view based sport video analysis using American football was demonstrated by Ding and Fan et al. (2002). Two generative models named the Hidden Markov Model (HMM) and the segmental HMM were used in for view based shot classification.

Automated analysis and summarization of soccer videos was described by Ekin et al. (2003) based on cinematic and object based features. It was composed of low level processing algorithm such as shot boundary detection, dominant color region detection and shot classification with higher level algorithms for goal detection, referee detection, and penalty box detection. Three categories of summary in a game: all slow motion segments, all goals and slow motion segments were obtained and classified using object based features. For speedy processing, the first two summaries were achieved based on cinematic features and the last type based on higher level semantics.

Taskiran et al. (2003) presented a framework for video detection using stochastic model. At first pre-processing was done wherein shot labels were obtained to determinate shot boundary in the given video sequence. After pre- processing, three types of features such as motion color and texture features were extracted from the each frame of the video sequence. The motion features were extracted by averaging horizontal and vertical components of the motion vector in the frame of the obtained shots. Similarly, the color features were obtained by averaging the pixel luminance and chrominance values within each frame and over the shot. Then, texture features were calculated by averaging the variance of pixel luminance values for each macro block within each frame and averaging these values for the shot. Then, GMM was modelled in order to obtain the number of cluster. Finally, HMM was adopted for video detection by exploiting the obtained cluster of GMM.

Spatial temporal feature extraction based video classification model was presented by Xu and Li (2003) for five TV broadcast video such as sports, cartoon, news, commercial and music. Two types of features- audio and video features were extracted from video frames. Audio features were extracted using MFCC (14 features) and visual features such as color, color layout, homogenous texture, mean and standard deviation were extracted from the video frames. On account of better classification audio and video features were concatenated. Then, PCA was undertaken to reduce the dimensional space of the combined features set into lower dimension and GMM classifier was utilized to classify the given video content into one the predefined category.

HMMs based sports video classification was presented by Gibert and Doermann (2003). For sports video classification, two types of features such as motion and color features were extracted. Motion features were extracted directly from the MPEG video streams. Color space of each video frame was taken as color features. Finally, two HMMs were employed by exploiting motion and color features separately. The output of two HMMs was integrated for final decision. Camera motion parameters (Takagi et al 2003) were used for sports video classification. Two camera based features: camera motion extraction ratio and camera motion transition which possess significant information for categorization of broadcasted sports video were considered.

Sports video summarization and its application to soccer game were demonstrated by Li et al. (2003). Automatic event detection in soccer game was implemented using automated analysis of visual and aural signals in the media. An ensemble classification approach for automated sports video classification system was described by Jaser et al (2003). It depends on the concept of cues which attach semantic meaning to low-level features computed on the video and audio. In order to classify sports category the multiple classifier system approach was adopted.

Edge based semantic classification of sports video sequences was described by Lee et al. (2003).It consists of two major stages: edge detection and semantic analysis. Generally, MPEG video has three basic frame types: Intra-coded (I), predictive coded (P), Bi-directional predictive coded (B) frames. Each I-frame was divided into 16x16 Macro Blocks (MBs) and each MB consists of four 8x8 luminance (Y) blocks and two 8x8 chrominance blocks. Each block was transformed into a DCT block, which

consists of one DC and 63 AC coefficients. On account of edge detection only AC coefficients of Y blocks in I-frames were used, whereas vertical and horizontal edge features were extracted. In semantic analysis video sequences were classified into predefined classes, whereas two step approaches were used. The first step was a coarse-level semantic classification based on edginess and the second step was a fine-level semantic classification based on duration.

Based on hierarchical decision, a scheme for sports video classification was demonstrated by Jaser et al. (2004). Pieces of visual evidence and characteristic of certain classes were extracted as key frames. Decision tree was used to analyse the semantics of the sports video content initially and then HMM was designed for categorization which was trained by the features extracted from the key frames.

Audio based sports video segmentation and detection was presented by Baillie and Jose (2004) for soccer video. At first, soccer sequence soundtrack was parameterized by using MFCC. Then, it was segmented into homogenous components by using windowing algorithm with Bayesian model based decision process. Finally, series of HMM classifier was used for soccer video classification.

The structure analysis of soccer video for classification and segmentation was explained by Xie et al. (2004) based on HMM .The salient features: dominant color ratio and motion intensity in compressed domain were extracted. Then, a set of HMM was modelled for each state. Color and motion features based sports news video shot classification was explained by Wang et al. (2004). Seven types of video shots and four types of sports video such as basketball, baseball, ice hockey and golf were examined. C4.5 decision tree classifier was employed for shot classification using the

extracted features. Edge feature based sports video classification was described by Yuan and Wan (2004). The popular canny edge operator was used to detect the edges in each video frame and KNN classifier was used for classification. To better exploit the edge information, edge detected image was dilated using morphological operation.

Vakkalanka et al. (2005) introduced a method for sports classification by combining multiple classifiers. Six kinds of TV broadcast video such as news, football, commercial, cartoon, tennis and cricket were considered for the sports classification. Spatial and temporal features were extracted from the video frames and fed into HMM and SVM classifier for classification. Similar work for the classification of four types of video clips such as news, outdoor, sports and drama was discussed by Gillespie and Nguyen (2005) using generalized and tree based radial basis function classifier.

A unified framework for sports video classification based on semantic shot classes was developed by Duan et al. (2005). It focused on clustering by key-frames with similar low level features. The framework was composed of four stages: Low level feature extraction, mid- level representation, shot attributes production, and classification. Initially, the low level features such as motion vectors field, texture map and pixel wise images in uncompressed domain were extracted from the given video sequence. Then nonparametric feature space analysis was adopted for mapping from low level features into mid -level representation such as motion, color and shot-length. Thirdly, feature vector for shot attributes description was constructed using obtained mid- level representations. Finally, SVM based supervised learning was adopted for classification of shot into one of the predefined shot categories.

Wu et al. (2005) presented an online leaning framework for sports video view classification using baseball video. An optimal local positive model was employed by sufficiently exploring the local statistic characteristics of the current under-test videos. A set of negative models were incorporated with the local positive model during the classification procedure to avoid adaptive threshold selection.

Semantic sports video sequence classification using hierarchical structure of audio-video was discussed by Kolekar and Sengupta (2005).Cricket video was taken into account for demonstration. At the top layer of hierarchical structure the input video was classified into event and non-event sequences, whereas audio features were used for classification. It was further classified into level-2 as spectators or on field. In level-3, the game actions were further sub-classified into real-time events and replay. At level-4, real-time shots were classified into wicket/hit departure and in level- 5 the wicket/hit class were further classified into wicket and hit. Finally in level- 6 post-activity was classified. HMM through dynamic programming using color or motion as a likelihood function was exploited from level 2 to level 6 classifications.

Automated classification of events in Australian Football League (AFL) was demonstrated by Tjondronegoro et al. (2005). The summarization approach starts with cinematic features which can be directly extracted from the raw video data. Based on camera-view, frames were classified into global, zoom-in and close up by exploiting experimental thresholds and domain knowledge. Based on camera-views classification, play-break sequences were segmented. In the end, the statistical characteristics of each play-break will be used to classify the sequence into one of AFL highlights, including goal, behind, mark, tackle and non-highlight.

The recognition of dynamic video contents was presented based on global probabilistic models by Piriou et al. (2006). At first the dynamic video contents were grouped into two categories named play segments and 'no play' using maximum likelihood criteria based on residual motion which accounts for the scene motion. Term weighting schemes plays an important role in text classification that can be applied to sports video classification. Five layered hierarchical framework based sports video classification was explained by Kolekar and Sengupta (2006). Top layer classification was performed by audio and video content analysis. Audio analysis was investigated by short time energy and zero crossing rate approaches. The video content was analysed by HMM by extracting color and motion features. Game specific rules were used by the lower layer to recognize major events of the game.

An automatic sports video classification system using sports video was described by Wang et al. (2006). On account of automatic reorganization of the sports video, multi-level framework was adopted. A pseudo 2-dimensional HMM classifier was used to classify the sports video by exploiting low-level visual/audio features. Hierarchical ontology based automatic video classification system was implemented by Yuan et al. (2006). Spatio-temporal features were extracted from the video sequence. On account of classification hierarchical SVM classifier was built by cross validation approach, which consists of a series of SVM classifiers united in a binary tree form.

Visual rhythm based low cost soccer video summarization was described by Bezerra and Lima (2006). It was applied in diverse tasks which aim at analysis the soccer videos such as shot transition detection, shot classification and attack direction

estimation. Play field color estimation approach was implemented initially by using HSV color space. To achieve color estimation of play field, the histogram of the Hue component of the VR (Vertical rhythm) image was computed. Similarly transition detection and scene classification was developed by exploiting k-means classifier on VR image. Camera direction detection was also employed to extract directional texture features from VR images.

Key frame based video summarization algorithm using Delaunay Clustering was described by Mundur et al. (2006).It consists of three steps: pre-sampling, clustering, and key frame selection. In the pre- processing stage, video frames were obtained from the original video. And color histograms were used to represent the visual content of video frames. Each frame was represented by a 256-bin color histogram in the HSV color space where there were 16 levels in H, 4 levels in S and 4 levels in V according to the MPEG-7 generic color histogram descriptor. Delaunay clustering approach was adopted for automated clustering which results in clusters of different sizes based on the content represented in the original video. Thus, key fames were obtained and video summarization was achieved.

Chasanis et al. (2007) presented a shot clustering approach for scene detection in video. At first, the given video was segmented into shots and key frames by global *k*-means clustering algorithm. Then, the segmented shots were clustered by applying spectral clustering approach based on visual similarity and the corresponding label was assigned to clustered shot.A statistical approach based shot segmentation and classification system was given by Yang et al. (2007). It segments and classifies the shots by using statistical inference features. Shot sequence probability was computed

by Bigram using the relations between adjacent shots and feature sequence probability which was dependent on inherent character of shot modelled by HMM.

A scheme of discriminant approach was designed for sports video classification using autocorrelogram by Watcharapinchai et al. (2007). The following sports video: boxing, golf, basketball, volleyball, tennis and football were considered for classification. Autocorrelogram robustly tolerates large changes in view positions of the camera zoom. On account of sports video classification two types of discriminant approaches: neural network with PCA and SVM were employed. Automated approach for personalized music sports video generation was described by Wang et al. (2007). Multimodal features were extracted from audio, video and text information. Dynamic programming was employed for the classification of personalized music sports video.

Bertini et al. (2007) introduced an automatic soccer video annotation approach using extended multimedia ontology. Enhanced HSV histogram with achromatic point detection based on a single hue and saturation parameter was used for video annotation. The more general concepts of the sport domain such as play, break and crowd were put in association with general visual features of the video like color and texture rather than domain specific visual features such as highlights, attack and action.

An approach for semantic scene classification in soccer video was presented by Kolekar and Palaniappan (2008). Shot detection was performed followed by classification based on clustering using shot aggregation. It was a top-down video scene classification approach composed with multiple level hierarchy processes. At

first, audio features were exploited to extract potentially interesting clips from the video. In the second level, clips were classified into field view and non-field view using feature of dominant grass color ratio. Then field view was categorized into three kinds of views using motion-mask and non-field view was classified into close-up and crowd using skin color information. In the final level close-up shots were classified into the four frequently occurring classes such as player of team *A*, player of team *B*, goal keeper of team *A*, goal keeper of team *B* using jersey color information. Finally, hierarchical classifier was adopted for semantic sports scene classification.

Two dimensional trajectories based video event classification and detection was presented by Hervieu et al. (2008). There were three dynamic video content understanding approaches employed by developing trajectory-based framework. The first one was clustering trajectories extracted from videos, whereas unsupervised solution was developed. Secondly events in videos were recognized. Thus, semantic classes of dynamic video contents were first learnt from a set of representative training trajectories. The third task was detecting unexpected events by comparing the test trajectory to representative trajectories of known classes of events by using HMM trajectory framework.

De Avila et al. (2008) presented a video summarization approach. At first, the given video was segmented into frames. Then, visual features were extracted from each frame by exploiting color histogram and line profiles (horizontal, vertical and diagonal) approaches. Next, the frames were grouped by an unsupervised clustering method, whereas the extracted visual features were used to identify frames with similar content. Finally, to create the static video summary, key frames were filtered

to eliminate frames that were too similar. In some cases, different key frames with very much similar visual content could be selected.

Li et al. (2009) presented a soccer video shot classification scheme based on dominant set clustering. Without any threshold setting, it automatically extracts dominant color region. Dominant sets clustering were adopted for dominant color distribution which naturally provides a principled measure of a cluster's cohesiveness as well as a measure of vertex participation to each group. The similarity between the dominant color regions of two frames was measured by the earth mover's distance, which was incorporated into the kernel function of SVM.

Li et al. (2009) presented an automated sports video and view type classification system using bag of visual-words model. At first, local significance for each frame undergoes Speeded Up Robust Features (SURF) extraction. After obtaining SURF model, a codebook was generated using *K* means clustering approach and each code word value was defined by the exemplar vector of each cluster. Then, the generated codebooks were used to calculate the Kullback-Leibler divergence between two probability distributions. Finally, KNN classifier was adopted for sports video classification based on Kullback-Leibler divergence.

Audio analysis based video classification was described by Rouvier et al. (2009) for cartoons, movies, news, commercials and music video. At first, perceptual linear predictive coefficients were extracted from the video frames as feature vector. Then, SVM classifier got trained by the extracted features and classified the unknown video into one of the predefined sports category while testing. The experiments were carried out using four types of video: cartoons, movies, news, commercials and music.

A framework for multimodal video classification using parallel neural networks was demonstrated by Montagnuolo and Messina (2009). Four categories of features were extracted from the given video. Initially, color, texture and motion features were extracted as visual perceptual information. Then structural information such as shot length, shot distribution, shot rhythm, shot clusters duration and saturation were obtained. Thirdly, cognitive information named face properties such as number, positions and dimensions were extracted followed by aural information such as transcribed text, sound characteristics extraction. Finally, the extracted features were used for parallel neural network training, which classified the unknown video into one the predefined video such as weather forecast, football, newscast, talk show, cartoons, music and commercials.

Halin et al. (2009) presented a shot view classification for play field based sports video. It was composed of two stages: play field segmentation and play field classification. At first, the input video was segmented into shots by applying ranking based approach and all frames within each shot were transformed into HSV color space from RGB color space. The play field segmentation was performed by three stages. Firstly HSV pixel-wise processing was taken place, whereas the hue component was assumed into logical within a particular chromatic range. At the same time, the achromatic components should also be constrained in order to avoid inclusion of pixels that were too dark or too bright. Then, noise removal and connected component analysis was adopted, whereas candidate playfield regions were marked as white pixels and non-playfield regions were marked as black pixels. Finally, the playfield-to-frame ratio was calculated and the play field region was classified into shot and close-up view.

A framework was implemented for automatic recognition and classification of scenes in TV program videos by Choros and Pawlaczyk (2010). To detect shots and then scenes in a tested TV program, automatic video indexer software was used. Edge features based sports video classification was demonstrated by Mohan and Yegnanarayana (2010). The features named edge direction histogram and edge intensity histogram were used to classify five kinds of sports video: volley ball, tennis, basketball, cricket and football using auto associative neural network models.

A framework for sports video classification was implemented by Sigari et al. (2011) based on ensemble classifier. Six features such as 3 dominant colors, cut rate, motion rate and dominant gray level were used as discriminant features. The following ensemble classifiers: decision tree, nearest neighbour, probabilistic neural network and linear discriminant analysis were used. The final decision was made by integrating the outputs of classifiers.

Semantic event detection and classification using key frame selection approach was introduced by Goyani et al. (2011) for cricket videos. It consists of different stages such as hue histogram difference based key frames selection for indexing, classification of frames into real time or replay based on logo transitions and also real time frames were classified into either field view or non-field view by color features.

Video content retrieval based sports video shot transition and classification was illustrated by Lijin (2011). An adaptive dual threshold was used for shot transition detection. Six common video shots, including sports news, venue shots, sports advertisements, volleyball games, soccer games, and table tennis games were

classified at the category level. Low level video features such as motion information, color and inter frame pixel differences were computed as feature vector for further classification. C4.5 decision tree classifier was employed for video shot classification.

Audio and visual analysis based automated consumer video summarization was introduced by Jiang et al. (2011). There were three approaches: overall quality of the key frames, quality of detected faces in the key frames, and the visual diversity of the selected key frames were developed for video summarization. The regression model was then applied to generate an aesthetic score roughly by measuring the image's quality. Then face quality was computed from detected faces by measuring the color contrast and lower score of the blur degree. To maintain the diversity of the selected key frames 5x5 grid-based color moments were extracted from the image frame, whereas large-enough distance frames were selected as key frames. These key frames were ranked in chronological order and were put together with the audio summary to generate the final video summary.

Bag of words approach for sports video classification was demonstrated by Duong et al. (2012). Initially, SURF was extracted from each frame. Then, a codebook was generated using SURF by employing k-means clustering algorithm. The histogram of codebook was computed and fed into SVM classifier for classification.

An automated sports video classification based on representative shot extraction and Geometry Visual Phrase (GVP) was presented by Dong et al. (2012). Initially, key frame clustering was adopted to select the shots which contain significant information of videos. Then, co-occurrence of visual words in a spatial

layout was used for GVP searching based on scale invariant feature transform. Finally, visual words and GVP were concatenated to form the enhanced histograms followed by SVM based classification.

SVM ensemble based video classification was demonstrated by Hamed et al. (2012). DCT coefficients were extracted from key frames in the given video sequences. Then PCA algorithm was adopted as dimensionality reduction approach. Finally, SVM ensemble classifier was used for video classification, whereas sports, news and movies were used. SVM based classification was employed by Capodiferro et al. (2012) to recover and preserve the historical sport videos of 1960 Olympic games. Features were extracted from the précised key frames in Laguerre Gauss transformed domain and were fed into SVM classifier for high level video classification. Mutchima and Sanguansat (2012) exploited term weighting scheme for video classification, whereas the given video was considered as a document while each frame was treated as words to identify the video contents. Color feature based sports video classification using HMM classifier was described by Hanna et al. (2012). Three categories of sports video such as hockey, football and golf were investigated. Color features were extracted from the three sports category and trained by HMM classifier for classification.

Ajmal et al. (2012) reviewed various video summarization approaches. They were feature based (motion, color, dynamic contents, gesture, audio-visual, speech transcript and object), cluster based (k-means, portioned clustering and spectral clustering), event based, shot based, trajectory based (spatial-temporal and curve simplification) and mosaic based approaches. Automated threshold and edge mapping

based video summarization and key frame selection was implemented by Dhagdi and Deshmukh (2012). Initially, the histogram difference of every frame was calculated, and then the edges of the candidate key frames were extracted by Prewitt operator. Finally, the edges of adjacent frames were matched. If the edge matching rate was above average edge matching rate, the current frame was deemed to the redundant key frame and should be discarded.

Approach for event detection was implemented by Xu et al. (2013) in soccer videos by creating audio keywords. Audio keyword is an effective and middle level representation that can link the gap between low level features and high level semantics. Low level audio features were used for audio keyword creation by exploiting SVM classifier. The created audio keywords can be used to detect semantic events in soccer video by applying a heuristic mapping. Mendi et al. (2013) presented a video summarization approach based on motion analysis for sports video. Two optical flow algorithms were used to estimate the motion metrics. Then, different key frame selection criteria's were used for each optical flow algorithms for video summarization and it was a threshold free approach.

Multi modal approach based user generated mobile sports video classification system was designed by Cricri et al. (2013). Visual features, MFCC features and auxiliary features were used for classification by SVM. Signature heat map based sports video classification was developed by Gade and Moeslund (2013). Thermal imaging was used for detecting players and homographs were used to detect their positions. The heat maps were generated by summarizing Gaussian distributions

representing people over 10-minutes periods. Before classification, low dimensional heat maps were produced using fisher faces.

A video categorization approach was implemented by Latt and War (2013) using Multivariate Adaptive Regression Splines (MARS). Five different video types: cartoons, sports, news, music and dahmas were analysed. MFCC3 and MFCC12 play a major role in MARS dahma models. For the MARS music models, MFCC3 and noise frame rate play major roles. Noise frame rate, silence ratio and spectral flux were considered as important features in MARS sports models. Short time energy and MFCC3 were important variable to decide the cartoon model. Noise frame ratio, short time energy, MFCC3 variables were more important in deciding the news models. Among these features MFCC3, MFCC12, short time energy, noise frame rate, silence ratio and spectral flux were considered as useful features for sports video classification.

Weighted Kernel Logistic Regression (WKLR) based video classification system was implemented by Hamed et al. (2013). Initially, key frames were selected from the given video sequences. Then, features were extracted by applying DCT on the selected key frames. As the size of DCT coefficients was same as input key frame, dimension reduction was taken into account. Hence, significant features were selected from the obtained DCT coefficient and insignificant features were discarded by exploiting PCA algorithm. Finally, video was classified by WKLR approach. The evaluations of WKLR approach were carried on three groups of video: sports, movies and news.

Ekenel and Semela (2013) introduced a multi modal video classification system for TV programs and YouTube videos. On account of video classification, various categories of descriptors were extracted. Visual descriptors represent color and texture based features and was used to describe the concepts appearing in a video. Wide range of perceptual cues was represented by the audio descriptors such as signal energy, zero crossing rate, fundamental frequency, and MFCC. Cognitive descriptors characterize the information derived from a face detector, whereas structural descriptors were related to shot editing of the video. Based on term frequency-inverse document frequency measure of YouTube videos, tag descriptor was used. Finally extracted features were fed into the SVM classifier for video classification.

Semantic based video indexing and video summarization was presented by Lei et al. (2014) based on key frame selection scheme. At first, video shots were divided into sub shots by dynamic distance reparability algorithm and semantic structure. Then, appropriate key frames were selected in each sub shot by singular value decomposition. Kapela et al. (2014) presented a real time sports scene classification system. Initially, video frames undergo preprocessing stage, whereas RGB to HSV color space conversion and color occurrence map was created. Then, 2-dimensional Fourier transform was adopted for frequency domain representation. The obtained Fourier representations were filtered with a set of Gabor filters. Finally, SVM with linear base function was adopted for sports scene classification.

Automatic video classification approach using multiple SVM notes was demonstrated by Jang et al. (2014). Five videos of videos such as animation, commercial, entertainment, drama, and sports were taken into account. Multiple SVM

classifiers were used that consider all possible binary grouping of the above mentioned sports groups. Given a query video, each SVM casts a probabilistic vote for each video. Finally, the optimal video was selected with the maximum votes.

An approach for video event classification was implemented by Chen and Tsai (2014) using multimodal features with HMM. Mid-level features provide clear linkages between low and high level audio visual features. For video event classification, it uses temporal context of mid-level multimodal features. In order to explore full temporal relations among multiple modalities in probabilistic HMM event classification, co-occurrence symbol transformation was adopted.

Spatiotemporal visual features based sports video classification was presented by Cricri et al. (2014). Three discriminative types of features such as visual, audio and auxiliary sensor data modalities were extracted for classification. Two categories of visual features such as spatial visual features and spatiotemporal visual features were extracted from player's movement and game flows. Audio track features were obtained using MFCC. Camera motion features were also extracted from video sequences, which were considered as an auxiliary sensor data. Finally, all these features were fused together using weighted fusion rule and fed into SVM classifier for sports video classification.

Higher order color moments based video summarization algorithm was presented by Jadhava and Jadhav (2015). It consists of two stages: shot boundary detection and key frame extraction. On account of video summarizing initially the input video was divided into multiple shots. Large changes in the video frame content can occur at the shot boundaries. Shots of video were detected using image histogram,

skewness and kurtosis. Also, key frames were extracted using the same algorithm, which was used for shot boundary detection.

Histogram clustering based video content retrieval approach was explained by Saravanan and Vengatesh (2015). The aim of clustering approach was grouping videos into various classes, which depends on the similarity parameter used to cluster the objects into different groups. Video clustering was performed by two searching process. The first searching process was on the image matrix and it was utilized to detect the centroids in order to remove the duplicate frames in the video. To create cluster, the second searching process uses image pixels and creates a novel matrix based indexing technique. Afterwards matrix cell histogram was calculated to retrieve the video from the video database. Kapela et al. (2015) presented a real time video classification approach. For feature computation Fast Fourier Transform (FFT) and Gabor filters were taken into account. Then decision tree and neural network classifiers were employed for sports event classification.

A detailed literature review for the classification of sports video was discussed in this chapter. It is noted from the literature review that the classification of sports video was mainly developed based on statistical and structural based features. Though there were many approaches, multi-resolution analysis has recently been proposed as a new method for feature extraction and image representation analysis. In order to explore and develop an automated and robust sports video classification framework, edge strength features are taken by exploiting NSST. In addition, multi class SVM classifier is employed for automated video classification. The following chapter discusses the mathematical background of NSST and SVM classifier.