Int J Multimed Info Retr (2012) 1:153–172
DOI 10.1007/s13735-012-0016-2

TRENDS AND SURVEYS

A study on video data mining

V. Vijayakumar · R. Nedunchezhian

Received: 17 October 2011 / Revised: 31 July 2012 / Accepted: 3 August 2012 / Published online: 25 August 2012

Abstract Data mining is a process of extracting previously unknown knowledge and detecting the interesting patterns from a massive set of data. Thanks to the extensive use of information technology and the recent developments in multimedia systems, the amount of multimedia data available to users has increased exponentially. Video is an example of multimedia data as it contains several kinds of data such as text, image, meta-data, visual and audio. It is widely used in many major potential applications like security and surveillance, entertainment, medicine, education programs and sports. The objective of video data mining is to discover and describe interesting patterns from the huge amount of video data as it is one of the core problem areas of the data-mining research community. Compared to the mining of other types of data, video data mining is still in its infancy. There are many challenging research problems existing with video mining. Beginning with an overview of the video data-mining literature, this paper concludes with the applications of video mining.

Keywords Video processing · Video information retrieval · Video mining · Video association mining · Movie classification · Sports mining

V. Vijayakumar (✉)
Research and Development Centre, Bharathiar University, Coimbatore 641 022, Tamil Nadu, India
e-mail: veluvijay20@gmail.com

V. Vijayakumar
Department of Computer Applications, Sri Ramakrishna Engineering College, Coimbatore 641 022, Tamil Nadu, India

R. Nedunchezhian
Department of Information and Technology, Sri Ramakrishna Engineering College, Coimbatore 641 022, Tamil Nadu, India
e-mail: rajuchezhian@gmail.com

1 Introduction

It is the advancement in multimedia acquisition and storage technology that has led to a tremendous growth in multimedia databases. Multimedia mining deals with the extraction of implicit knowledge, multimedia data relationships or other patterns not explicitly stored in the multimedia data [6,45,100]. The management of multimedia data is one of the crucial tasks in the data mining owing to the non-structured nature of the multimedia data. The main challenge is to handle the multimedia data with a complex structure such as images, multimedia text, video and audio data [21,60,65,72].

Nowadays people have accessibility to a tremendous amount of video both on television and internet. So, there is a great potential for video-based applications in many areas including security and surveillance, personal entertainment, medicine, sports, news video, educational programs and movies and so on. Video data contains several kinds of data such as video, audio and text [59]. The video consists of a sequence of images with some temporal information. The audio consists of speech, music and various special sounds whereas the textual information represents its linguistic form.

The video content may be classified into three categories, namely [102]. (i) Low-level feature information that includes features such as color, texture, shape and so on, (ii) Syntactic information that describes the contents of video, including salient objects, their spatial-temporal position and spatial-temporal relations between them, and (iii) semantic information, which describes what is happening in the video along with what is perceived by the users. The semantic information used to identify the video events has two important aspects [87]. They are: (a) A spatial aspect presented by a video frame, such as the location, characters and objects displayed in the video frame. (b) A temporal aspect presented by a

sequence of video frames in time such as the character's actions and the object's movements presented in a sequence. The higher-level semantic information of video is extracted by examining the features of the audio, video and text of the video. Multiple cues from different modalities including audio and visual features are fully exploited and used to capture the semantic structure of the video bridging the gap between the high level semantic concepts and the low-level features. Three modalities are identified within a video [84]. They are: (1) Visual modality containing the scene that can be seen in the video; (2) Auditory modality with the speech, music, and environmental sounds that can be heard along with the video; (3) Textual modality having the textual resources which describes the content of the video document.

Video databases are widespread and video data sets are extremely large. There are tools for managing and searching within such collections, but the need for tools to extract the hidden and useful knowledge embedded within the video data is becoming critical for many decision-making applications.

1.1 Video processing

Though the acquisition and storage of video data is an easy task the retrieval of information from the video data is a challenging task. One of the most important steps involved is to transform the video data from non-structured data into a structured data set as the processing of the video data with image processing or computer vision techniques demands structured-format features [73,75]. Before applying the data-mining techniques on the video key frame, the video, audio and text features are extracted using the image processing techniques, eliminating the digitalization noise and illumination changes to avoid false positive detection.

The video data can be structured in two ways according to the content structure. First, and foremost the scripted video databases [105,111] are carefully produced according to a script or plan that is later edited, compiled and distributed for consumption. These video databases have some content structures such as movies, dramas and news. Second, unscripted video databases [105] have no content structures as "raw" videos like surveillance videos and sports videos have no scene change.

It is the most fundamental task in video processing to partition the long video sequences into a number of shots and find a key frame of each shot for further video information retrieval tasks. Hu et al. [40] presented several strategies in visual content-based video indexing and retrieval to focusing on the video structure analysis, including shot boundary detection, key frame extraction and scene segmentation, extraction of features and video data mining. Bhatt and Kankanhalli [6] discussed the video feature extraction, video transformation and representation techniques and the video data-mining techniques.

1.1.1 Video data model

Since, the relational or object oriented data model does not provide enough facilities for managing and retrieving the video contents efficiently, an appropriate data model is needed to do it. Three main reasons [79] can be identified for this: (1) lack of facilities for the management of spatiotemporal relations, (2) lack of knowledge-based methods for interpreting raw data into semantic contents and (3) lack of query representations for complex structures.

A video data model is a representation of video data based on its characteristics and content as well as the applications it is intended for [44]. It is based on the idea of video segmentation or video annotation. Video data mining requires a good data model for video representation. Various models have been proposed by different authors. Petkovic and Jonker [79] proposed a content-based retrieval data model with four layers. They are: (i) Raw video data layer with a sequence of frames, as well as some video attributes. (ii) Feature layer consisting of domain-independent features that can be automatically extracted from raw data, characterizing colors, textures, shapes, and motion. (iii) Object layer having the entities, characterized by a prominent spatial dimension and assigned to regions across frames. (iv) Event layer with entities that have a prominent temporal extent describing the movements and interactions of different objects in a spatial-temporal manner.

Zhu et al. [111] described a hierarchical video database management framework using video semantic units to construct database indices. They presented a hierarchical video database model [27] that captures the structures and semantics of video contents in databases. It provides a framework for automatic mapping from the high-level concepts to the low-level representative features. It is exploited by partitioning the video contents into a set of hierarchically manageable units such as clusters, sub clusters, sub regions, shots or objects, frames or video object planes and regions. It supports a more efficient video representation, indexing and video data accessing techniques.

1.1.2 Video segmentation

The first step in any video data management system is invariably, the segmentation of the video track into smaller units [3,68] enabling the subsequent processing operations on video shots, such as video indexing, semantic representation or tracking of the selected video information and identifying the frames where a transition takes place from one shot to another. The visual-based segmentation identifies the shot boundaries and the motion-based segmentation identi-

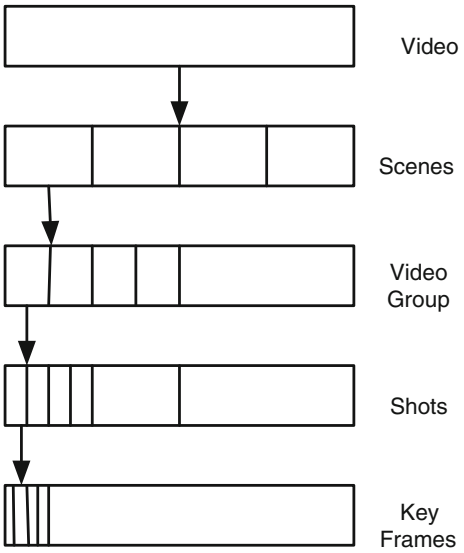

Fig. 1 Video hierarchy

fies pans and zooms [12]. In general, most of the videos from daily life can be represented using a hierarchy of levels [29] shown in Fig. 1. The following terms are defined [68,105] as

Video It refers to multimedia sequences comprised of both sound and a series of images.

Scene It is a collection of semantically related and temporally adjacent groups depicting and conveying a high-level concept with a sequence of shots focusing on the same point or location of interest [83].

Video group It is an intermediate entity between the physical shots and semantic scenes serving as the bridge between the shot and scene. Two kinds of shots are absorbed into a video group [116]. (1) Temporally related: Shots related in temporal series where similar shots are shown back and forth. (2) Spatially related: Shots similar in visual perception, where all shots in the group are similar in visual features.

Shot It is defined as a sequence of frames taken by a single camera with no major changes in the visual content [82]. One of the important initial steps in segmentation and analysis of the video data is the shot-boundary detection. What is challenging in video segmentation is that the shot change detection algorithm must be able to handle all types of scene changes like abrupt changes and gradual changes.

Key frame The frame represents the salient visual contents of a shot. Since a large number of adjacent frames are likely to be similar, one or more key frames can be extracted from the shot depending on the complexity of the content of the shot. The key frames extracted from the video streams are treated as images.

1.1.3 Feature extraction

The video segmented and the key frames chosen, the low-level image features can be extracted from the key frames. The low-level visual features such as color, texture, edge and shapes can be extracted and represented as feature descriptors [43]. A feature is defined as a descriptive parameter extracted from an image or a video stream [92]. There are two kinds of video features extracted from the video [11,45]. (i) The description based features that use metadata, such as keywords, caption, size and time of the creation. (ii) The content based features are based on the content of the object itself. There are two categories of content based features: the global features extracted from a whole image and the local or regional features describing the chosen patches of a given image [36]. Each region is then processed to extract a set of features characterizing the visual properties including the color, texture, motion and structure of the region. The shot-based features and the object-based features are the two approaches used to access the video sources in the database.

Fu et al. [29] extracted the video features such as color histogram, domain color, edge color and texture and stored to XML documents for further video mining process. Choudhary et al. [18] represented the low-level features like the position of objects, size of objects, color correlogram of objects, and so on extracted from the frames of the video or high level semantically meaningful concepts like the object category and the object trajectory [18].

The spatial features (visual), audio features and the features of moving objects (temporal features) can be used to mine the significant patterns in a video [48]. These resulting patterns are then evaluated and interpreted to obtain the knowledge of the application. The Mining objects, events and scenes in the video are useful for many applications. For instance, they can serve as entry points for retrieval and browsing or they can provide a basis for video summarization [80]. Liu and He [56] detected football events by using the multiple feature extraction (visual, auditory features, text and audio keywords) and fusion.

1.2 Video information retrieval

A video retrieval system is essentially a database management system. It is for handling video data is concerning with returning similar video clips (or scenes, shots, and frames) to a user for their video query. For an information retrieval to the efficient, the video data have to be manipulated properly. Basically there are four steps involved in any automatic video information retrieval [75,73,83], namely (1) Shot boundary detection, (2) Key frames selection, (3) Extracting low-level features from key frames and (4) Content-based video information retrieval with a query in the form of input provided

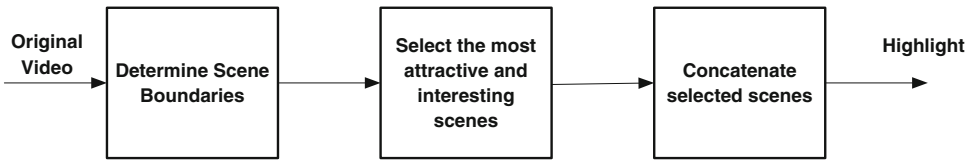

Fig. 2 A video highlight

by the user and a search carried out through the database on the basis of the input.

There are the two familiar concepts used for the video information retrieval: video abstraction and video annotation.

1.2.1 Video abstraction

Video Abstraction is a short summary of a long video having a set of stills or moving images selected and reconstructed from an original video with concise information about the content widely used in the video cataloging, indexing and retrieving [73,75]. The video abstraction is of two types: video summarization and dynamic video skimming.

Video Summarization Video Summarization is defined as a sequence of stills or moving pictures (with or without audio) presenting the content of a video [83] in such a way that the respective target group is rapidly provided with concise information about the content preserving the essential message of the original and aiming at generating a series of visual contents for users to browse and understand the whole story of the video efficiently [10,51]. Selecting or reconstructing a set of salient key frames from an original video sequence, it discards similar frames preserving the frames, different from one another and drawing the attention of the people. It assumes that the scripted content summary is carefully structured as a sequence of semantic units whereas the unscripted content requires a "highlights" extraction framework that captures only remarkable events that constitute the summary.

The Video summarization uses two approaches [107] based on the video features. First, the rule-based approach combines evidences from several types of processing (audio, video, text) to detect certain configuration of events included in the summary using the pattern recognition algorithms to identify elements in the video and finds rules to qualitatively select important elements to be included in the summary. Ngo et al. [70] proposed a unified approach for summarization based on the analysis of the video structures and video highlights. Second, the mathematically oriented approach that uses similarities within the videos to compute a relevant value of video segments or frames with the mathematical criteria, such as frequency of occurrence to quantitatively evaluate the importance of the video segments. Lu et al. [51] presented an approach for video summarization based on the graph optimization. The approach has three stages: First, the source video is segmented into video shots selecting a candidate shot set from the video shots according to some video features. Second, a dissimilarity function is defined between the video shots to describe their spatial-temporal relation and the candidate video shot set is modeled into a directional graph. Third, a dynamic programming algorithm is used to search the longest path in the graph as the final video skimming generating a static video simultaneously.

Dynamic Video Skimming Video skimming [73,75] consists of a collection of image sequences along with the related audios from an original video possessing a higher level of semantic meaning or preview of an original video than the video summary does and highlight generation is one of the important applications in it with the most interesting and attractive parts of a video [75]. It is similar to a trailer of a movie showing the most attractive scenes without revealing the ending of a film and used in a film domain frequently as shown in Fig. 2.

1.2.2 Video annotation

Annotation involves attaching keywords from the specialism along with commentaries produced by experts and those obtained from a body of the existing texts [85]. The text embedded in the video (closed caption) is a powerful keyword resource in building the video annotation and retrieval system enabling text-based querying, video retrieval and content summarization.

Video annotation [67] has preceded along three dimensions, namely, supervised annotation (uses machine learning of visual and text features), unsupervised annotation and contexts-based approaches. In supervised learning based on the annotated concepts for each video, the unclassified files are automatically categorized. In unsupervised learning, the video files are clustered and annotators assign key words to each cluster which can be used to extract rules for annotating future documents. The third approach tries to mine concepts by looking at the contextual information such as the text associated with images to derive semantic concepts. Moxley et al. [67] presented a new approach to automatic video annotation by leveraging search and mining techniques. It employed as a two-step process of search followed by mining. Given a query video consisting of the visual content and speech-recognized transcripts, the similar videos are first ranked in a multimodal search. Then, the transcripts associated with these similar videos are mined to extract keywords for the query. Tseng et al. [95] proposed an innovative method for

the semantic video annotation by integrating visual features, speech features and frequent patterns existing in a video. Wang et al. [101] presented an approach to automatically annotate a video shot with an adaptive number of annotation key words according to the richness of the video content.

Annotation [95] can be classified into three categories: (a) Statistics-Based Annotation that computes the probabilities between visual features and candidate keywords; (b) Rule-Based Annotation that discovers the associated concepts hidden in the sequential images; (c) Hybrid Annotation Methods which integrate the statistics-based and rule-based methods for mining visual features in the video. The annotation problem has received a significant attention, since annotation helps bridge the semantic gap that results from querying using one mode (e.g., text) for the returns of another mode (e.g., images) [57]. Generating captions or annotations automatically for still video is an arduous task. Aradhye et al. [5] presented a method for large scale auto-annotation videos without requiring any explicit manual annotation.

This paper presents a detailed study on video data-mining concepts, techniques, research issues and various application domains are discussed here. The paper is organized as follows. Section 2 discusses the concepts of video data mining. Section 3 presents the video data-mining techniques and approaches. Section 4 explains the application of the video data mining. Section 5 covers the key accomplishments of video data mining. The research issues and future directions are discussed in Sect. 6 and Sect. 7 concludes the paper.

2 Video data mining

It is video data mining that deals with the extraction of implicit knowledge, video data relationships, or other patterns not explicitly stored in the video databases considered as an extension of still image mining by including mining of temporal image sequences [64]. It is a process which not only automatically extracts content and structure of video, features of moving objects, spatial or temporal correlations of those features, but also discovers patterns of video structure, object activities, video events from vast amounts of video data with a little assumption of their contents.

2.1 Video information retrieval versus video data mining

It is video information retrieval, not information retrieval that is considered as part of the video data mining of a certain level knowledge discovery such as feature selection, dimensionality reduction and concept discovery.

The dissimilarities of video data mining [101] with related areas are as follows,

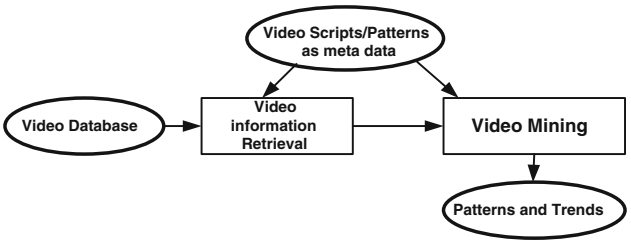

Fig. 3 Video mining

- Video data mining versus computer vision or video processing: The relationship between video processing and video mining is very subjective. The goal of video data mining is to extract patterns from the video sequences whereas video processing focuses on understanding and/or extracting features from the video database.
- Video data mining versus pattern recognition: Both areas share the feature extraction steps but differ in pattern specificity. The objective of pattern recognition is to recognize specific, classification patterns, pattern generation and analysis. Pattern recognition is indulging in a research on classifying special samples with an existing model while video mining is involved in discovering rules and patterns of samples with or without image processing. The objective of video data mining is to generate all significant patterns without prior knowledge of what they are.
- Video information retrieval versus video data mining [75,73]: The difference is similar to the difference between database management and data mining [94]. Video mining focuses on finding correlations and patterns previously unknown from large video databases shown in Fig. 3.

Sanjeevkumar and Praveenkumar [86] presented an architecture for multimedia data mining to find the patterns in multimedia data. Video mining involves three main tasks [23]. They are: (1) Video preprocessing with high quality video objects such as blocks of pixels, key frames, segments, scenes, moving objects and description text; (2) The extracting of the features and semantic information of video objects such as physical features, motion features, relation features and semantic descriptions of these features, and (3) Video patterns and knowledge discovery using video, audio and text features.

2.2 Key problems in video data mining

Video data mining is an emerging field that can be defined as the unsupervised discovery of patterns in audio visual contents. Mining video data is even more complicated

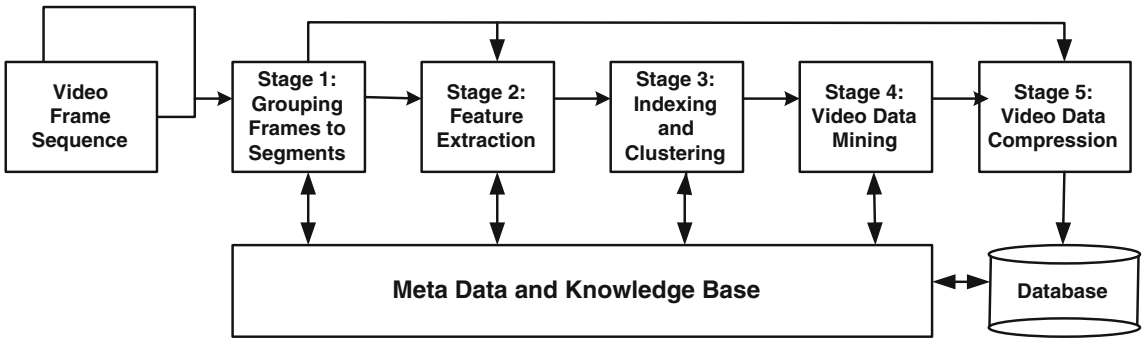

Fig. 4 General framework for video data mining

than mining still image data requiring tools for discovering relationships between objects or segments within the video components, such as classifying video images based on their contents, extracting patterns in sound, categorizing speech and music, and recognizing and tracking objects in video streams. The existing data-mining tools pose various problems while applied to video database. They are: (a) Database model problem in which video documents are generally unstructured in semantics and cannot be represented easily via the relational data model demanding a good video database model that is crucial to support more efficient video database management and mining [111]. To adopt a good model, one needs to address three problems, namely (1) How many levels should be included in the model? (2) What kind of decision rules should be used at each node? (3) Do these nodes make sense to human beings? (b) The retrieval results solely based on the low level feature extraction are mostly unsatisfactory and unpredictable. It is the semantic gap between the low level visual features and the high level user domain that happens to the one of the hurdles for the development of a video data-mining system. Modelling the high level features rather than the low level features is difficult as the former is depending on the semantics whereas the latter is based on the syntactic structure. (c) Maintaining data integrity and security in video database management structure. These challenges have led to a lot of research and development in the area of video data mining. The main objective of video mining is to extract the significant objects, characters and scenes by determining their frequency of re-occurrence.

Most of the existing video mining tools lack the semantic interpretation of the video. Though there are several widely-accepted data-mining techniques, most of them are unsuitable for video data mining because of the semantic gap. Numerous methodologies have been developed and many applications have been investigated including the organizing video data indexing and retrieval [111,115] extracting representative features from raw video data before the mining process and integrating features obtained from multiple modalities. But still, it is in need of improved methods for the retrieval and mining process. Some of the current directions in mining video data include [93] extracting data and/or metadata from the video databases, storing the extracted data in structured databases, and applying data-mining tools to the structured video databases, integrating data-mining techniques with the information retrieval tools and developing data-mining tools to operate directly on the unstructured video databases.

2.3 Video data mining

Video mining can be defined as the unsupervised discovery of patterns in an audio-visual content [23]. The temporal (motion) and spatial (color, texture, shapes and text regions) features of the video can be used for the task mining.

Oh and Bandi [72] proposed a framework for real time video data mining for the raw videos which is shown in Fig. 4. In the first stage the grouping of input frames takes place to a set of basic units. In the second stage it extracts some of the features from each segment. In the third stage, the decomposed segments are clustered into similar groups. The next two are the actual mining of the raw video sequences and the video data compression for the storage of these raw videos. The knowledge and patterns can discover and detect the object identification, modeling and detection of normal and abnormal events, video summarization, classification and retrieval.

Oh and Bandi [72] and Su et al. [91] proposed a multi-level hierarchical clustering approach to group segments with similar categories at the top level and similar motions at the bottom level using K-Means algorithm and cluster validity method.

2.4 Audio data mining

Audio is what plays a significant role in the detection and recognition of events in video. Supplying speech, music and various special sounds and can be used to separate different speeches, detect various audio events, analyze for spoken

text, emotions, high-light detection in sports videos and so on. In movies, the audio is often correlated with the scene [82]. For instance, the shots of fighting and explosions are mostly accompanied by a sudden change in the audio level.

In addition, audio features can be used to characterize the media signals to discriminate between music and speech classes. In general, the audio features [86] can be categorized into two groups, namely, time domain features which include zero-crossing rates, amplitudes, pitches and the frequency domain features consisting of spectrograms, cepstral coefficients and mel-frequency cepstral coefficients. There are two main approaches to audio data mining [53]. Firstly, Text-based indexing approach converts speech to text and then identifies words in a dictionary having several hundred thousand entries. If a word or name is not in the dictionary, the Large Vocabulary Continuous Speech Recognizers system will choose the most similar word it can find. Secondly, phoneme-based indexing approach analyzes and identifies sounds in a piece of audio content to create a phonetic-based index. It then uses a dictionary of several dozen phonemes to convert a user's search term to the correct phoneme string. Finally, the system looks for the search terms in the index.

Zhou et al. [110] proposed the data-mining models for detecting errors in Dictation Speech Recognition. They presented three popular data-mining techniques for detecting errors, including Naive Bayes, Neural Networks and Support Vector Machines. Doudpota and Guha [25] proposed a system to automatically locate and extract songs from digitized movies. Subsequently a song grammar was proposed and used to construct a probabilistic timed automaton to differentiate songs. Audio classification is being used in various applications such as musical instrument sound identification, music retrieval or personal recognition. Okada et al. [76] introduced the rule-based classification method for multi-class audio data. Audio power spectra of multiclasses are transformed into a transaction database that includes a "class label item" in each transaction. Classification rules are extracted by an exhaustive search of the closed itemsets and the greedy rule-selection approach. Chen et al. [13] proposed an unsupervised technique of discovering commercials by mining repeated sequence in audio stream.

2.5 Video text mining

It is the video texts which highlight the important events in the video database such as the speech transcriptions and sports overlay/sliding texts that are the information source and exploiting several information extraction techniques to arrive at the representative semantic information. Being useful in the labeling of the videos [63] the video text can be obtained from three sources: scene text, superimposed text and automatic speech recognition [59]. Scene text occurs as a natural part of the actual scene captured by the camera. The examples are, billboards, text on vehicles, and writings on human clothes; Super imposed text is mechanically added text to the video frames in order to supplement the visual and audio content providing additional information for a better understanding of the video and Automatic speech recognition converts speech to text is then mined. Nemrava et al. [69] presented the semantic multimedia annotation and indexing with the use of the video textual resources. Kucuk and Yacici [46] proposed a text-based fully automated system for the semantic annotation and retrieval of the news videos exploiting a wide range of information extraction techniques including the named entity recognition, automatic hyper-linking, person entity extraction with co-reference resolution and semantic event extraction.

3 Video data mining approaches

Recently, there has been a trend of employing various data-mining approaches [61,72,91] in exploring knowledge from the video database. Consequently, many video mining approaches have been proposed which can be roughly classified into five categories. They are: Video pattern mining, Video clustering and classification, Video association mining, Video content structure mining and Video motion mining.

3.1 Video structure mining

Since, video data is a kind of unstructured stream an efficient access to video is not an easy task. Therefore the main objective of the video structure mining is the identification of the content structure and patterns to carry out the fast random access of the video database.

As video structure represents the syntactic level composition of the video content, its basic structure [29] is represented as a hierarchical structure constituted by the video program, scene, shot and key-frame as shown in Fig. 1. Video structure mining is defined as the process of discovering the fundamental logic structure from the preprocessed video program adopting data-mining method such as classification, clustering and association rule.

It is essential to analyze video content semantically and fuse multi-modality information to bridge the gap between human semantic concepts and computer low-level features from both the video sequences and audio streams. Video structure mining gets not only the video content constructing patterns but also the semantic information among the constructive patterns [20]. Video structure mining is executed in the following steps [115]: (1) video shot detection, (2) scene detection, (3) scene clustering, and (4) event mining.

Fu et al. [29] defined two kinds of structural knowledge, namely, video semantic and syntactic structure knowledge

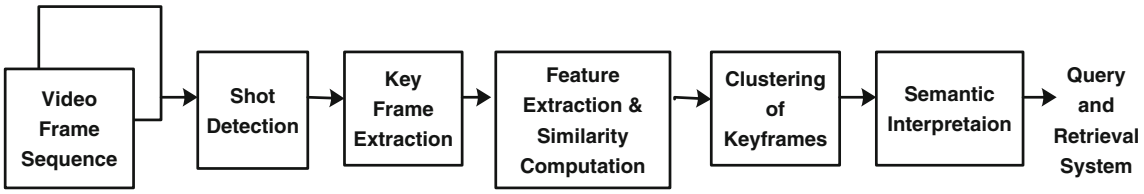

Fig. 5 Video clustering

leading to the concepts of video semantic and syntactic structure mining. Syntactic structure mining is based on the video basic structure which adopts the methods of data mining according to the similar video units and video unit features. It acquires some syntactic rules in general, including dialogue, interview, news, talk show and so on. These video syntactic rules are structural knowledge triggering the process that mines constructional patterns in the video structure units, and explores relations between video units and features. Semantic structure mining is a process that discovers semantics and events in video basic structure units. The basic structure units explore the relations between video unit features and features such as color and texture pattern in the explosion scene, light and texture pattern in indoor or outdoor scene, audio pattern in highlight scene and so on. These relations are represented by association rules between video unit feature(s) and feature(s).

The current researches on it focus on mining object semantic information and event detection. The video event represents the occurrences of certain semantic concepts. Chen et al. [12,14,15] presented a video event detection framework that is shot-based, following the three-level architecture and proceeding the low-level descriptor extraction, mid-level descriptor extraction, and high-level analysis. Heuristic rules can be used to partially bridge the semantic gap between the low-level features and the high level subjective concepts. The decision tree logic data classification model algorithm is then performed upon the combination of multimodal mid-level descriptors and the low-level feature descriptors for event detection. Zhao et al. [114] proposed the Hierarchical Markov Model Mediator mechanism to efficiently store, organize, and manage the low-level features, multimedia objects, and semantic events along with the high-level user perceptions such as user preferences in the multimedia database management system.

3.2 Video clustering and classification

Video clustering and classification are used to cluster and classify video units into different categories. Therefore clustering is a significant unsupervised learning technique for the discovery of certain knowledge from a dataset. Clustering video sequences in order to infer and extract activities from a single video stream is an extremely important problem and so it has a significant potential in video indexing, surveillance,activity discovery and event recognition [97,103]. In the video surveillance systems, it is to find the patterns and groups of moving objects that the clustering analysis is used. Clustering similar shots into one unit eliminates redundancy and as a result, produces a more concise video content summary [116,117]. Clustering algorithms are categorized into partitioning methods, hierarchical methods, density-based methods, grid based methods and model-based methods.

Vailaya et al. [99] proposed a method to cluster the video shots based on the key frames representing the shots. Figure 5 shows a block diagram of the general problem of video clustering. It is detecting by the shots from the video frame sequence that the key frames are extracted. Next, a feature vector is computed so that the key frames can be clustered based on the feature vector assigning the semantic interpretations to various clusters at the last stage. These semantic interpretations are used in the retrieval system to index and browse the video database. At the clustering stage, it is desirable to cluster the shots into semantic categories such as presence/absence of buildings, texts, specific texture and so on, so that a higher level abstract (semantic) label can be assigned to the shots (indoor shots, outdoor shots, city shots, beach shots, landscape shots).

Video classification aims at grouping videos together with similar contents and to disjoin videos with non-similar contents and thus categorizing or assigning class labels to a pattern set under the supervision. It is the primary step for retrieval and the classification approaches are those techniques that split the video into predefined categories. Semantic video classification approaches [26] can be classified into two categories. First, rule-based approach that uses domain knowledge to define the perceptional rules and achieve semantic video classification and easy to insert, delete and modify the existing rules when the nature of the video classes changes. It is attractive only for the video domains such as news and films that have well-defined story structures for the semantic units (i.e., film and news making rules). Second, the statistical approach that uses statistical machine learning to bridge the semantic gap. This supports more effective semantic video classification by discovering non-obvious correlations (i.e., hidden rules) among different video patterns.

The key features are used to categorize video into predefined genres. Video classifications are based on the spatial and temporal characteristics and necessary for efficient access, understanding and retrieval of the videos. Pan et al. [77] proposed a video graph tool for video mining and visualizing the structure of the plot of a video sequence. The video graph of the video clip is the directed graph where every node corresponds to a shot group and edges indicate temporal succession. This algorithm is used to "stitch" together similar scenes even if they are not consecutive and automatically derive video graphs. It derives the number of recurrent shot groups for video mining and classification, distinguishing between different video types, e.g., news stories versus commercials.

Pan et al. [78] presented a video cube tool to classify a video clip into one of 'n' given classes (e.g., "news", "commercials", etc) which automatically derive a "vocabulary" from each class of the video clips, using the "Independent Component Analysis" incorporating the spatial and temporal information which works on both video and audio information. It creates a vocabulary that describes images, motions and the audio parts of the video and thus provides a way to automatically extract features. The video and audio features reveal the essential characteristics of a genre class and are closely related to the neural signals used in the human perceptual press. VCube algorithm uses the video bases of genre classes to classify a video clip and the audio bases to classify the clips based on their audio information.

Building an activity recognition and classification system is a challenging task because of the variations in the environment, objects and actions. Variations in the environment can be caused by cluttered or moving background, camera motion, occlusion, weather and illumination changes while Variations in the objects are because of the differences in appearance, size or posture of the objects or because of self motion which is not a part of the activity and variations in the action can make it difficult to recognize semantically equivalent actions as such, for example imagine the many ways to jump over an obstacle or different ways to throw a stick.

Nowozin et al. [71] proposed a classifier for the sequence representations for the action classification in the videos that retains the temporal order in a video. They first proposed the LPBoost classifier for sequential representations, and then, the discriminative PrefixSpan subsequence mining algorithm to find the optimal discriminative subsequent patterns. Brezeale et al. [7] came out with a survey on video classification. They found that features are drawn from three modalities divided into four groups of automatic classification of the video such as text-based approaches, audio based approaches, visual-based approaches, and the combination of the text, audio, and visual features. Tien et al. [96] extracted the high-level audiovisual features to describe the video segments which are further transformed to symbolic streams and an efficient mining technique was applied to derive all frequent patterns that characterize tennis events. After mining, they categorized the frequent patterns into several kinds of events and thus achieved event detection for tennis videos by checking the correspondence between mined patterns and events.

3.3 Video association mining

Video association mining is the process of discovering associations in a given video. The video knowledge is explored in a two stages, the first being the video content processing in which the video clip is segmented into certain analysis units extracting their representative features and the second being the video association mining that extracts the knowledge from the feature descriptors. Mongy et al. [66] presented a framework for video usage mining to generate user profiles on a video search engine in the context of movie production that analyzes the user behaviors on a set of video data to create suitable tools to help people in browsing and searching a large amount of video data.

In video association mining, the video processing and the existing data-mining algorithms are seamlessly integrated into mine video knowledge. Zhu et al. [111] proposed a multilevel sequential association mining to explore the associations between the audio and visual cues and classified the associations by assigning each of them with a class label using their appearances in the video to construct video indices. They integrated the traditional association measures (support and confidence) and the video temporal information to evaluate video associations.

Sivaselvan et al. [89] presented a video association mining consisting of two key phases. First, the transformation phase converts the original input video into an alternate transactional format, namely, a cluster sequence. Second the frequent temporal pattern mining phase that is concerned with the generation of the patterns subject to the temporal distance and support thresholds.

Lin et al. [55] developed a video semantic concept discovery framework that utilizes multimodal content analysis and association rule mining technique to discover the semantic concepts from video data. The framework used the apriori algorithm and association rule mining to find the frequent item-sets in the feature data set and generated the classification rules to classify the video shots into different concepts (semantics). Chen and Shyu [14] proposed a hierarchical temporal association mining approach that integrates the association rule mining and the sequential pattern discovery to systematically determine the temporal patterns for target events. Goyani et al. [33] proposed an A-priori algorithm to detect the semantic concepts from the cricket video. Initially, a top-down event detection and classification was performed using the hierarchical tree. Then the higher

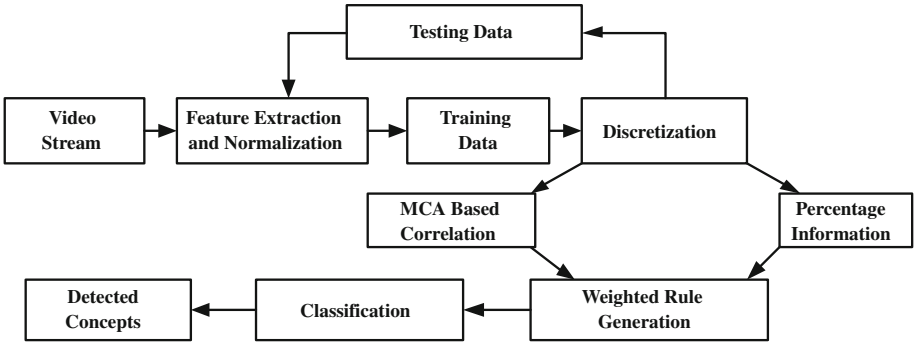

Fig. 6 Semantic concept detection

level concept was identified by applying A-Priori algorithm. Maheshkumar [58] proposed a method that automatically extracts silent events from the video and classifies each event sequence into a concept by sequential association mining. A hierarchical framework was used for soccer (football) video event sequence detection and classification. The association for the events of each excitement clip was computed using an a priori mining algorithm using the sequential association distance to classify the association of the excitement clip into semantic concepts.

Lin and Shyu [52] proposed weighted association rule mining algorithm able to capturing the different significant degrees of the items (feature-value pairs) and generating the association rules for video semantic concept detection shown in Fig. 6. The framework first applies multiple correspondence analyses to project the features and classes into a new principle component space and discovers the correlation between feature-value pairs and classes. Next, it considered both correlation and percentage information as the measurement to weight the feature-value pairs and generate the association rules. Finally, it performs classification by using these weighted association rules.

Kea et al. [47] developed a method based on the frequent pattern tree (FPTree) for mining association rules in video retrieval. The proposed temporal frequent pattern tree growth algorithm mine temporal frequent patterns from TFPTree for finding the rules of the motion events.

3.4 Video motion mining

Motion is a key feature that essentially characterizes the contents of the video, representing the temporal information of videos and more objective and consistent compared to other features such as color, texture and so on. There have been some approaches to extract camera motion and motion activity in video sequences. While dealing with the problem of object tracking, algorithms are always proposed on the basis of known object region in the frames and so the most challenging problem in the visual information retrieval is the recognition and detection of the objects in the moving videos. The camera motion having a vital role to play some of the key issues in video motion detections are, the camera placed in static location while the objects are moving (surveillance video, sports video); the camera is moving with moving objects (movie); multiple cameras are recording the same objects. The camera motion itself contains a copious knowledge related to the action of the whole match. The important types of camera motion are Pan (left and right), Zoom (in and out), Tilt (up and down), and Unknown (camera motions those are not Pan, Zoom, or Tilt are grouped to Unknown).

Wu et al. [104] proposed the extraction scheme of the global motion and object trajectory in a video shot for content-based video retrieval. For instance, while browsing the video obtained by surveillance system or watching sports programs, the user always has the need to find out the object moving in some special direction. Zang and Klette [112] proposed an approach for extraction of a (new) moving object from the background and tracking of a moving object.

Mining patterns from the movements of moving objects is called motion mining. First, the features are extracted (physical, visual and aural, motion features) using objects detection and tracking algorithms and then the significations of the features, trends of moving object activities and patterns of events are mined by computing association relations and spatial-temporal relations among the features.

3.5 Video pattern mining

Video pattern mining detects the special patterns modeled in advance and usually characterized as video events such as dialogue, or presentation events in medical video. The existing work can be divided into two categories such as mining similar motion patterns and mining similar objects [4].

Sivic et al. [90] described a method for obtaining the principal objects, characters and scenes in a video by measuring the reoccurrence of the spatial configurations of the viewpoint invariant features has three stages: The first stage extracts the neighborhoods occurring in more than a minimum number of key frames considered for clustering, where as the second stage matches the significant neighborhoods using a greedy progressive clustering algorithm, and in the

third stage, the resulting clusters are merged based both on spatial and temporal overlap. Burl et al. [8] presented an algorithm to extract information from raw, surveillance-style video of an outdoor scene containing a mix of people, bicycles, and motorized vehicles. A feature extraction algorithm based on the background estimation and subtraction followed by spatial clustering and multi-object tracking was used to process sequences of video frames into a track set, which encodes the positions, velocities, and the appearances of the various objects as the function of time are mined to answer the user-generated queries. Lai et al. [50] proposed a motion model that enables to measure the similarities among different animal movements in high precision. A clustering method can separate the recurring movements from the infrequent random movements.

Fleischman et al. [28] presented an approach in which the temporal information is captured by representing events using a lexicon of hierarchical patterns of human movement that are mined from a large corpora of un-annotated video data. These patterns are used as features for a discriminative model of event classification that exploits tree kernels in a Support Vector Machine. The second category systems aim at grouping frequently appearing objects in videos. Therefore, it is useful to have commonly occurring objects/characters/scenes for various applications [90]. There is a number of applications: First, they provide entry points for visual search in video databases. Second, they can be used in forming video summaries—the basic elements of a summary often involve the commonly occurring objects and these are then displayed as a storyboard. The third application area is in detecting product placements in a film where the frequently occurring logos or labels are prominent. Mining repeated short clips from video collections and streams are essential for video syntactic segmentation, television broadcast monitoring, commercial skipping, content summary and personalization, as well as video redundancy detection and many other applications.

Xie and Chang [106] investigated the pattern mining strategies in video streams. They applied different pattern mining models (deterministic and statistic; static and temporal) and devised pattern combination strategies for generating a rich set of pattern hypothesis. Some of the statistical clustering method such as K-means, HMM, HHMM and Deterministic algorithms were considered for video clustering. Yang et al. [108] proposed a method to repeat the clip mining and the knowledge discovery from the video data. The mining framework unifies to detect both the unknown video repeats and the known video clips of the arbitrary length by the same feature extraction and matching process. The structure analysis method is effective in discovering and modeling the syntactic structure of the news videos and their main objective is to detect the unknown video repeats from the video stream without prior knowledge. Su et al. [91] presented a method to achieve an effective content-based video retrieval by mining the temporal patterns in the video contents. It was the construction of a special index on video temporal patterns for an efficient retrieval (Fast-Pattern- Index tree) and a unique search strategy for effective retrieval (Pattern-based Search).

4 Video data mining applications

The fact that video data are used in many different areas such as sports, medicine, traffic and education programs, shows how significant it is. The potential applications of video mining include annotation, search, mining of traffic information, event detection / anomaly detection in a surveillance video, pattern or trend analysis and detection. There are four types of videos [63,72] in our daily life, namely, (a) produced video, (b) raw video, (c) medical video, and (d) broadcast or prerecorded video.

4.1 Produced video data mining

A produced video is meticulously produced according to a script or plan that is later edited, compiled and distributed for consumption. News videos, dramas, and movies are examples of the produced video with an extremely strong structure but has tremendous intra-genre variation in production styles that vary from country to country or content-creator to content-creator. The success and significance of the video mining depends on the content genre [23,24].

For news videos detection of story boundaries either by closed caption and/or speech transcript analysis or by using speaker segmentation and face information have been proved effective whereas for movie contents, the detection of syntactic structures like two-speaker dialogues and also detection of specific events like explosions have been proved immensely useful and for situation comedies, the detection of physical setting using mosaic representation of a scene and the detection of the major cast using audio-visual cues have also been beneficial.

Shirahama et al. [88] focused on the rhythm in a movie, consisting of the durations of the target character's appearance and disappearance. Based on this rhythm, they divided the movie into topics, each topic corresponding to one meaningful episode of the character. By investigating such topics, they discovered immensely useful editing patterns of character's rhythm supported by their semantic features. These rhythms can also be used to annotate certain topics. Shirahama et al. [87] proposed a video data-mining approach with temporal constraint for extracting the previously unpredictable semantic patterns from a movie. First, having transformed a movie of an unstructured raw material into a multi-stream of raw level metadata, they extracted the

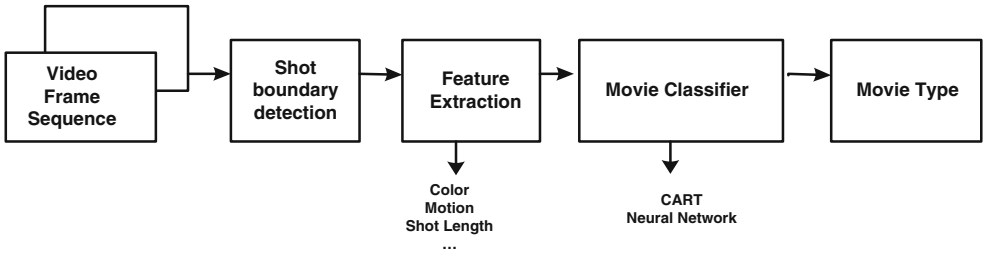

Fig. 7 Movie classification

sequential patterns then from the multi-stream of raw level metadata using a parallel data mining.

Rasheed et al. [81] proposed a method to classify movies into four broad categories such as Comedies, Action, Dramas and Horror Films. The video features such as average shot length, color variance, motion content and lighting key are combined in a framework to provide a mapping to these four high-level semantic classes.

Huangy et al. [35] presented the film classification method that consists of three steps as shown in Fig. 7. Boundary detection being the first step, they analyzed the color, motion and brightness from every shot, and represented the shot with these low-level features in the second step. The final step is the classification process, as there are many classification methods, like K-mean, classification tree, mean-shift, ada-boost, neural network and so on. Generally, these methods can distinguish between the supervised and un-supervised classes. In the future, the classification result can be improved by combining audio or text cues.

There is a lot of editing patterns available in video editing which is a process of selecting and joining various shots to create final video sequence. Discovering the editing patterns, the video material is edited to precisely convey the editor's intention to a viewer by using a universal rule called video grammar.

Matsuo et al. [61] proposed the methods to automatically extract editing rules specific to each video by using the data-mining technique. Two approaches are used to detect the editing patterns in a video stream. They are (i) The extraction of patterns from a multi-symbol stream, and (ii) the extraction of periodic patterns in time-series data.

4.2 Raw video data mining

There are two common types of surveillance video used in the real world applications such as the security video generally used for property or public areas and the monitoring video used to monitor the traffic flow. The surveillance systems with data-mining techniques are investigated to find out suspicious people capable of indulging in abnormal activities. However, the captured video data are commonly stored or previewed by operators to find abnormal moving objects or events. The identification of the patterns existing in surveillance applications, building the supervised models and the abnormal event detection are risky tasks [22]. The semantic events are the incidents captured by the surveillance video on the road, such as car crash, bumping, U-turn and speeding.

Chen et al. [17] proposed a video data-mining framework that analyzes the traffic video sequences by using background subtraction, image/video segmentation, object tracking, and modeling with multimedia augmented transition network model and multimedia input strings, in the domain of traffic monitoring over an intersection. Dai et al. [20,22] proposed a surveillance video data-mining framework of motion data that discovers the similar video segments from surveillance video through a probabilistic model. A mixture of hidden Markov models using an expectation-maximization scheme is fitted to the motion data to identify the similar segments. These surveillance systems with data mining techniques are being investigated to find out suspicious people capable of carrying out terrorist activities.

The collected raw video data in the traffic databases applications cannot provide organized, unsupervised, conveniently accessible and easy-to-use multimedia information to the traffic planners. The analysis and mining of traffic video sequences to discover information such as vehicle identification, traffic flow and the spatio-temporal relations of the vehicles at intersections provides an economic approach for daily traffic operations. In order to discover and provide some important but previously unknown knowledge from the traffic video sequences for the traffic planners, video data-mining techniques need to be employed.

Choudhary et al. [18] proposed a framework for automated analysis of the surveillance videos using cluster algebra to mine various combinations of patterns from the component summaries (such as time, size, shape, position of objects, etc.) to learn the usual patterns of events and discover unusual ones. Praveenkumar et al. [49] proposed a framework to discriminate between normal and abnormal event in a surveillance video as shown in Fig. 8. In the framework, the audio-visual features are extracted from the incoming data stream and the resultant real valued feature data is binarized. A feature selection process based on association rule mining

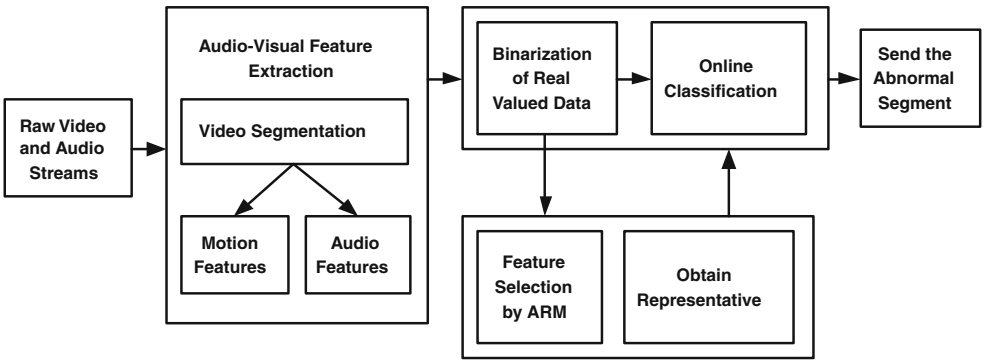

Fig. 8 Abnormal detection in a surveillance video

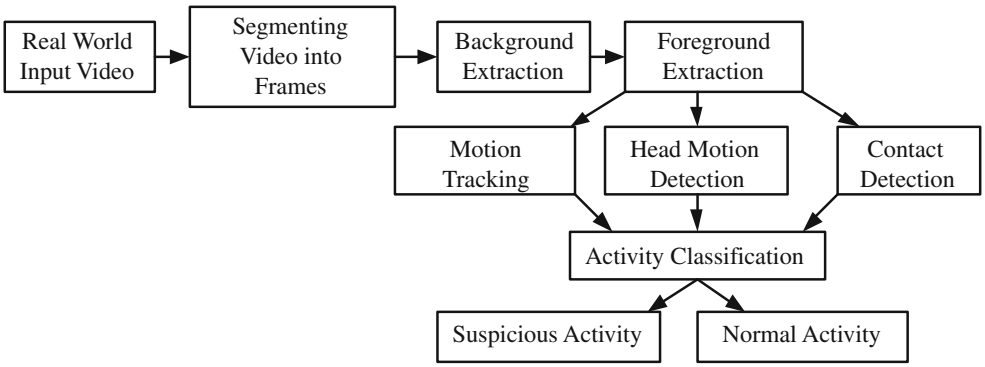

Fig. 9 Human activity recognition system

has selected highly discriminant features. A short representative signature of the whole database is generated using a novel reservoir sampling algorithm stored in binary form and used with a Support Vector Classifier to help discriminate events as normal or abnormal event.

Jiang et al. [74] proposed a context-aware method to detect anomalies. By tracking all the moving objects in the video, three different levels of spatiotemporal contexts are considered, i.e., point anomaly of a video object, sequential anomaly of an object trajectory, and co-occurrence anomaly of multiple video objects. A hierarchical data-mining approach was proposed. At each level, frequency-based analysis was performed to automatically discover regular rules of normal events and thus events deviating from these rules are identified as anomalies.

Gowsikhaa et al. [30] proposed a method to detect suspicious activities such as object exchange, entry of new person, peeping into other's answer sheet and person exchange from the video captured by a surveillance camera during examinations based on the face and hand recognition. Figure 9 illustrates the brief design of Human Activity Recognition system.

Anwar et al. [1] presented a framework to discover the unknown anomalous events associated with a frequent sequence of events that is to discover events, unlikely to follow a frequent sequence of events in surveillance videos.

4.3 Medical video mining

Audio and video processing is integrated to mine the medical event information such as dialog, presentation and clinical operation from the detected scenes in a medical video database.

Zhu et al. [116] developed a video content structure and event-mining framework to achieve more efficient video indexing and access for medical videos. Visual and audio feature processing techniques are utilized to detect some semantic cues such as slides, face and speaker changes, within the video and these detection results are put together to mine three types of events (presentation, dialogue, clinical operation) from the detected video scenes. Zhang et al. [109] presented an intra-vital video mining system of leukocytes rolling and adhesion. Video mining of vivo microscopy video sequences is very difficult because of severe noises, background movements, leukocytes deformations, and contrast changes. Aligning the video frames to eliminate the noises caused by camera movements, they located the moving leukocytes by applying and comparing a spatiotemporal probabilistic learning method and a neural network framework for time series data. Finally, they removed the noises by applying the median and location-based filtering. They proposed a new method for the automatic recognition of the non-adherent and adherent leukocytes.

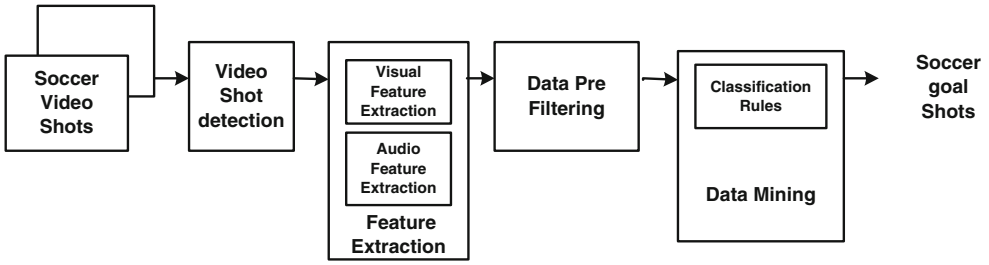

Fig. 10 Sports video mining

4.4 Broadcast or prerecorded video mining

Broadcast video can be regarded as being made up of genre (set of video documents sharing similar style). The genre of a video is the broad class to which it may belong to e.g. sports, news and cartoon and so on. The content of broadcast video can be conceptually divided into two parts. First, the semantic content, the story line told by the video. This is split into genre, events and objects. Second, inherent properties of the digital media video termed as editing effects.

In broadcast video data (unscripted content), such as sports video and meetings video, the events happen spontaneously. Data mining can be used by sports organizations in the form of statistical analysis and pattern discovery as well as outcome prediction. Although sports videos (non-edited) are considered as non-scripted, they usually have a relatively well-defined structure (such as the field scene) or repetitive patterns (such as a certain play type) helping us enhances the scriptedness of sports videos for more versatile and flexible access. A sports game usually occurs in one specific playfield but it is often recorded by a number of cameras with fixed positions also.

In sports video data, Mosaic is generated for each shot as the representative image of the shot content [62] such a mosaic based approach provides two kinds of mining methods: unsupervised mining of the structure without prior knowledge, and supervised mining of key-events with domain knowledge. Without prior knowledge, the play is mined by unsupervised clustering on mosaics as well as a voting process. With prior knowledge, the key-events are mined using Hidden Markov Models.

Chen et al. [16] proposed sports video mining framework (Fig. 10) and first analyzed the soccer videos by using joint visual and audio features. Then the data pre-filtering step was performed on raw video features with the aid of domain knowledge and classification rules were used to extract the goal events. It can be used for the high-level indexing and selective browsing of soccer videos.

Tien et al. [94] proposed a mining-based method to achieve event detection for broadcasting tennis videos. Initially extracting some high-level features to describe video segments, they further transformed the audiovisual features into symbolic streams and an efficient Max-Sub pattern Tree mining technique to derive all frequent patterns that characterize tennis events. After mining, they categorized frequent patterns into several kinds of events and thus achieved event detection for tennis videos by checking the correspondence between mined patterns and events.

Ding and Fan [19] presented a multichannel segmental hidden Markov model for sports video mining that incorporates ideas from Coupled HMM, segmental HMM, and hierarchical HMMs. This model offers more flexibility, functionality and capacity than its precedents in two aspects.

Harikrishna et al. [36] presented an approach for classification of events in cricket videos and summarized its visual content. Using sequential pattern mining and support vector machine they classified the sequence of shots into four events, namely, RUN, FOUR, SIX and OUT.

5 Key accomplishments of video mining

Video mining techniques are used for automatically generating critical decisions in numerous current real world problems. Video data mining provides the critical data needed to verify strategies for product placement, product assortment and cross merchandising. The key problem addressed includes view-independent person detection, multi-person tracking, a method for specifying behaviors, and robust behavior recognition. The approach is to use a variety of computer vision and statistical learning techniques under the constraints of a typical retail environment. The extremely crowded scenes pose unique challenges to video analysis that cannot be addressed with conventional approaches. The key insight is to exploit the dense activity of the crowded scene by modeling the rich motion patterns in local areas, effectively capturing the underlying intrinsic structure they form in the video.

Video semantic event detection has become more and more attractive in recent years such as video surveillance [113], sports highlights detection [36], TV/Movie abstraction and home video retrieval [31] and so on. Mining the television programs for getting narrative style time and effect patterns of advertisements. Guha et al. [34] proposed the surveillance event characterization at two levels. First, a set

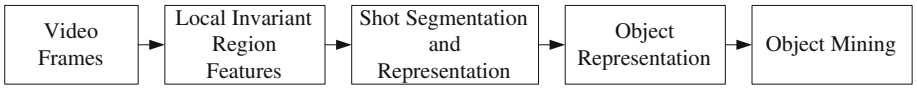

Fig. 11 Framework for object mining

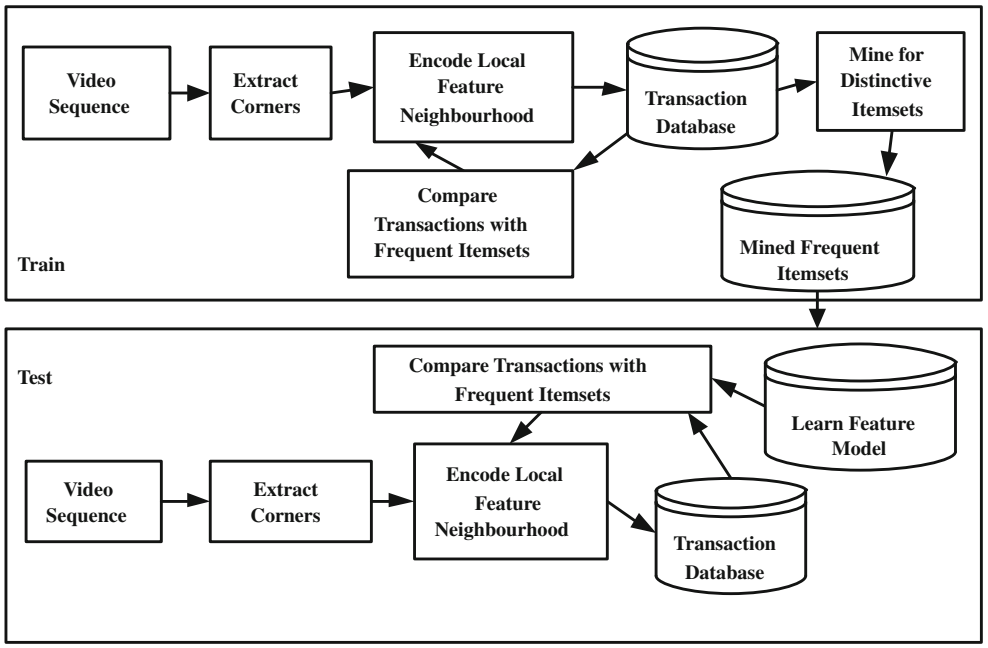

Fig. 12 Recognizing abnormal actions within video sequences

of time-varying predicates defined on heterogeneous objects moving in unknown environments. Second, the information characterizable is used to index and retrieve the event characterizations.

The mining and automatic analysis surveillance video extracts a valuable knowledge. The examples of knowledge and patterns that can discover and detect from a surveillance video sequence are object identification, object movement pattern recognition, spatio-temporal relations of objects, modeling and detection of normal and abnormal (interesting) events, and event pattern recognition [74].

The applications of surveillance mining are detecting crowd patterns for traffic control, recognition of suspicious people in large crowds, detecting traffic patterns, monitoring of surveillance in theft or fire protection and care of bedridden patients and young children.

Anjulan and Canagarajah [2] proposed a framework for performing object mining where video segmentation, feature extraction, tracking, clustering, object retrieval and mining are combined seamlessly within a framework as shown in Fig. 11. Initially, video is segmented into a number of smaller shots, and then, the features are extracted from the shots to represent the visual information. Next, these features are grouped into clusters to represent relevant objects, each cluster approximately corresponding to an object in a shot. These object clusters may contain a number of similar instances of the same object, and these instances are grouped together in the mining stage.

Oh et al. [74] developed a model to capture the location of motions occurring in a segment using the two dimensional matrix. The segmented pieces being clustered using the features, an algorithm is used to find whether a segment has normal or abnormal events by clustering and modeling normal events. Lee et al. [54] proposed a model-based conceptual clustering of moving objects in video surveillance the basis of on formal concept analysis. The formal concept analysis utilized to generate concepts, handle complicated moving objects and provides conceptual descriptions of moving object databases such as significant features and relationships. Movie data serve as a good test case of mining for knowledge consisting of some structure such as shots and scenes more than one media and structural boundaries. The knowledge can be mined for the movie such as composite structure analysis, identifying interesting events and patterns, clustering the movies, cinematic rules and rhymes of characteristic topics.

Gilbert et al. [32] presented an approach for recognizing actions within movie video sequences as shown in Fig. 12. Initially, 2D corners are detected in three orthogonal planes of the video sequence. Each corner encoded as a three-digit number denoting the spatiotemporal plane. They are used within an iterative hierarchical grouping process to form descriptive compound features. Each corner is grouped within a cuboid-based neighborhood. A set of grouped corners is called a Transaction and these are collected to form a Transaction database which is then mined for finding the most

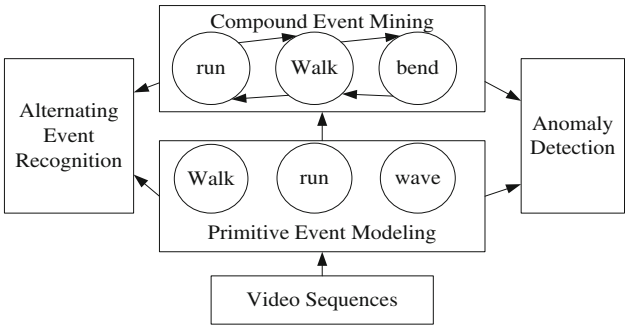

Fig. 13 Framework for hierarchical visual event pattern mining

frequently occurring patterns. These patterns are descriptive, distinctive sets of corners, and are called frequent item sets which then become the basic features for mining.

Cui et al. [9] proposed a hierarchical visual event pattern mining framework and its applications on recognition and anomaly (both abnormal events and abnormal contexts) detection, including an approach to unsupervised primitive event categorization, an extended Sequential Monte Carlo method for primitive and compound event recognition, and a method for abnormal events and contexts detection as shown in Fig. 13.

In the future, video mining will continue to receive attention, especially for its application in online video sharing, security surveillance monitoring and effective image retrieval. Anwar et al. [42] presented a framework to discover the unknown anomalous events associated with a frequent sequence of events.

Some research organizations are contributing much in the area of video mining such as, DIMACS Workshop [37], MERL [39] and DVMM lab of Columbia University [38]. However, only a limited work has been done in video data mining. Table 1 shows the summary of the video data-mining key activities.

6 Research issues in video data mining

There are five main research issues for video mining: video semantic event detection and concept mining [52], video object motion analysis [2], representation of video data model [111], extracting the appropriate video features, and selecting the video data-mining algorithms.

The mining of semantic concepts in video data is a core challenge of modern video data management, opening applications like intelligent content filters, surveillance, personal video recommendation, or content-based advertisement. The main challenge is the semantic gap problem which focuses on how to predict semantic features from primitive features. There should be a general framework that is not domain specific like news, sports and so on that can detect semantic features from the general videos that are applicable for any kind of videos. There is a need for a system that can extract, analyze and model the multiple semantics from the videos by using the primitive features.

Semantic event detection is still an ordinary problem owing to the large semantic gap and the difficulty of modelling temporal and multimodality characteristics of video

Table 1 Summary of video data-mining key activities

Authors	Year	Concepts	Application domain
Gowsikhaa et al. [30], Anwar [1]	2012	Temporal association rule mining, suspicious human activity detection	Security and surveillance
Cui et al. [9]	2011	Hierarchical visual event pattern mining and anomaly detection	Surveillance
Gilbert et al. [32] Harikrishna et al. [36], Jiang et al. [42]	2011	Sequential pattern mining and support vector machine, hierarchical data mining,	Movie, sports, surveillance
Anjulan and Canagarajah [2] Gaidon et al. [31], Ding and Fan [19]	2009	Object mining, support vector machine, multichannel segmental hidden markov model	Movie/TV shows, sports video
Tien et al. [94], Huangy et al. [35], Choudhary et al. [18], Praveenkumar et al. [49]	2008	Symbolic streams mining, visual feature mining, association rule mining and support vector machine	Sports, movie, surveillance
Lee et al. [54], Zhang et al. [109]	2007	Model-based conceptual clustering, automatic spatiotemporal mining	Medical
Guha et al. [34], Dai et al. [20,22]	2006	Multi-agent Tracking, surveillance video data mining	Surveillance
Rasheed et al. [81]	2005	Classification	Movie
Shirahama et al. [87,88]	2004	Parallel data mining, movie editing pattern	Movie
Oh et al. [74] Chen et al. [16]	2003	Normal or abnormal event segmentation by clustering, multimedia data mining framework	Sports video
Zhang [113], Zhu et al. [116], Matsuo et al. [61]	2002	Independent motion detection, event detection, video editing patterns	Surveillance, medical, movie

streams. The temporal information plays a vital role in the video data mining particularly, in mining and recognizing patterns in film, medical, sports and traffic videos. For example, in the basketball games the history of situations in terms of time-series of states is more vital than distinct states/processes or actions/events to find the correct zone-defense strategy detection. It is indeed useful to know that not only the current positions of each defender, but also their previous positions and movements are a matter of concern.

The analysis of motion characteristics for moving objects in video is an important part of video data mining. For example, in sports one needs to analyze the behaviors and technical features of athletes by their motion trajectory. In the future it is intended to explore the case of fast moving trajectory tracking and multiple moving objects mining algorithms. 3D motion analysis and object based video annotation also considered to improve the performance mining process.

Video data mining needs a model selection because of the different domains having different descriptive complexities [111]. So the need to develop the general framework to all video domains and evaluate the performance of the video mining algorithm in environments containing more events is of paramount importance. The purpose of the generative video model to bridge the semantic gap between the high-level concepts and the low-level/mid-level features.

Extracting optimal features is an ongoing major challenge. Optimal features should be tailored to the specific application such as motion tracking or event detection, and also utilize multimodal aspects (audio, visual, and text).

Selecting a proper video mining algorithm is a challenging issue. It needs a little assumption for video data. It handles different types and lengths of video data, but it should also represent the patterns with effective model. It must interpret and use the mined semantic information and patterns and soft computing techniques should also be studied in order to improve the results.

Video data mining may be beneficial to (i) reduce the dimension space for storage saving and computation reduction; (ii) advance learning methods to accurately identify target semantics for bridging the semantics between low-level/mid-level features and high-level semantics; (iii) effectively search media content for dynamical media delivery and enable the extensive applications to be media-type driven; (iv) customizable framework for video indexing that can index the video by using the modalities according to the user preferences.

7 Conclusion

In this paper, a brief overview of video data mining is presented. With over a decade of extensive research, there has been a tremendous development and application activities in the video data-mining domain. It is impossible to give a complete coverage on this topic with limited space and knowledge. There are many challenging research problems facing video mining such as discovering knowledge from spatial-temporal data, inferring high-level semantic concepts from the low-level features extracted from videos and making use of unlabeled data. The detection of unusual and abnormal video events is indispensable for consumer video applications such as sports highlights extraction and commercial message detection as well as surveillance applications. To improve the results of the video data mining, the new features can be constructed by analyzing the heterogeneous data like video text, audio, and videos. Besides spatial features, there are temporal features, audio features, and features of moving objects in the video data and all these features can be used to mine. There is no meaningful clustering or segmentation method that can be universally applied to all kinds of visual media. However, in-depth research is still required on several critical issues so that there can be developments in leaps and bounds in the data-mining field.

References

1. Anwar F, Petrounias I, Morris T, Kodogiannis V (2012) Mining anomalous events against frequent sequences in surveillance videos from commercial environments. Exp Syst Appl 39:4511–4531
2. Anjulan A, Canagarajah N (2009) A unified framework for object retrieval and mining. IEEE Trans Circ Syst Video Technol 19(1):63–76
3. Ahmed A (2009) Video representation and processing for multimedia data mining. Semantic mining technologies for multimedia databases. IGI Press, pp 1–31
4. Anjulan A, Canagarajah N (2007) A novel video mining system. In: Proceedings of 14th IEEE international conference on image processing, San Antonio, Texas, pp 185–189
5. Aradhye H, Toderici G, Yagnik J (2009) Video2Text: learning to annotate video content. In: Proceedings of IEEE international conference on data mining workshops, pp 144–152
6. Bhatt CA, Kankanhalli MS (2011) Multimedia data mining: state of the art and challenges. Multimedia Tools Appl 51:35–76
7. Brezeale D, Cook DJ (2008) Automatic video classification: a survey of the literature. IEEE Trans Syst Man Cybern Part C: Appl Rev 38(3):416–430
8. Burl MC (2004) Mining Patterns of activity from video data. In: Proceedings of the SIAM international conference on discrete mathematics, pp 532–536
9. Cui P, Liu Z-Q, Sun L-F, Yang S-Q (2011) Hierarchical visual event pattern mining and its applications. J Data Mining Knowl Disc 22(3):467–492
10. Chen B-W, Wang J-C, Wang F (2009) A novel video summarization based on mining the story-structure and semantic relations among concept entities. IEEE Trans Multimedia 11(2):295–313
11. Colantonio S, Salvetti O, Tampucci M (2008) An infrastructure for mining medical multimedia data. Lect Notes Comput Sci 5077:102–113
12. Chen F, Cooper M, Addock (2007) Video summarization preserving synamic content. In: Proceedings of the international workshop on TRECVID video summarization, pp 40–44

13. Chen J, Li T, Zhu L, Ding P, Xu B (2011) Commercial detection by mining maximal repeated sequence in audio stream. Proceedings of IEEE
14. Chen M, Chen S-C, Shyu M-L (2007) Hierarchical temporal association mining for video event detection in video databases. In: The second IEEE international workshop on multimedia databases and data management (MDDM'07), in conjunction with IEEE international conference on data engineering (ICDE2007), Istanbul, Turkey
15. Chen S-C, Chen M, Zhang C, Shyu M-L (2006) Exciting event detection using multi-level multimodal descriptors and data classification. In: Proceedings of eighth IEEE international symposium on multimedia, pp 193–200
16. Chen S-C, Shyu M-L, Zhang C, Luo L, Chen M (2003) Detection of soccer goal shots using joint multimedia features and classification rules. In: Proceedings of international workshop on multimedia data mining (MDM/KDD'2003), USA, pp 36–44
17. Chen S-C, Shyu M-L, Zhang C, Strickrott J (2001) Multimedia data mining for traffic video sequences. In: Proceedings second international workshop on multimedia data mining MDM/KDD'2001 in conjunction with ACM SIGKDD seventh international conference on knowledge discovery and data mining, pp 78–86
18. Choudhary A, Chaudhury S, Basnerjee S (2008) A framework for analysis of surveillance videos. In: Proceedings of sixth Indian conference on computer vision, graphics & image processing, pp 344–350
19. Ding Y, Fan G (2009) Sports video mining via multi-channel segmental hidden Markov models. IEEE Trans Multimedia 11(7):1301–1309
20. Dai K, Zhang J, Li G (2006) Video mining: concepts, approaches and applications. Proc IEEE 2006:477–481
21. Djeraba C (2003) Multimedia mining: a highway to intelligent multimedia documents. Springer, Berlin
22. Dai KX, Li GH, Gan YL (2006) A probabilistic model for surveillance video mining. In: Proceedings of the fifth international conference on machine learning and, cybernetics, pp 1144–1148
23. Divakaan A, Peker K, Chang S, Radhakrishnan R, Xie L (2004) VideoMining: pattern discovery versus pattern recognition. In: Proceedings IEEE international conference on image processing (ICIP'2004). Mitsubishi Electric Research Laboratories
24. Divakaran A, Miyahara K, Peker KA, Radhakrishnan R, Xiong Z (2004) Video mining using combinations of unsupervised and supervised learning techniques. In: Proceedings of SPIE conference on storage and retrieval for multimedia databases, vol 5307, pp 235–243
25. Doudpota SM, Guha S (2011) Mining movies to extract song sequences. In: Proceedings of MDMKDD'11
26. Fan J, Luo H, Elmagarmid AK (2004) Concept-oriented indexing of video databases: towards semantic sensitive retrieval and browsing. IEEE Trans Image Process 13(7):974–992
27. Fan J, Zhu X, Hacid M-S, Elmagarmid AK (2002) Multimedia tools and applications. Model-based video classification toward hierarchical representation indexing and access. Kluwer, Dordrecht, pp 97–120
28. Fleischman M, Decamp P, Roy D (2006) Mining temporal patterns of movement for video content classification. In: Proceedings of the 8th ACM international workshop on multimedia, information retrieval, pp 183–192
29. Fu C-J, Li G-H, Dai K-X (2005) A framework for video structure mining. In: Proceedings of the fourth international conference on machine learning and cybernetics, vol 3, pp 1524–1528
30. Gowsikhaa D, Manjunath AS (2012) Suspicious human activity detection from surveillance videos. Int J Int Distrib Comput Syst 2(2):141–149
31. Gaidon A, Marszalek M, Schmid C (2009) Mining visual actions from movies. In: Proceedings of the British machine conference. BMVA Press, pp 125.1–125.11
32. Gilbert A, Illingworth J, Bowden R (2011) Action recognition using mined hierarchical compound features. IEEE Trans Pattern Anal Mach Intell 33(5):883–897
33. Goyanil M, Dutta S, Gohil G, Naik S (2011) Wicket fall concept mining from cricket video using a-priori algorithm. Proc Int J Multimedia Appl (IJMA) 3:1
34. Guha P, Biswas A, Mukerjee A, Sateesh P, Venkatesh KS (2006) Surveillance video mining. In: Proceedings of the third international conference on visual information engineering, Bangalore (India), September 26–28, 2006
35. Huangy H-Y, Shih W-S, Hsu W-H (2008) A film classifier based on low-level visual features. J Multimedia 3(3):26–33
36. Harikrishna N, Satheesh S, Dinesh Sriram S, Easwarakumar KS (2011) Temporal classification of events in cricket videos. In: Proceedings of seventeenth national conference on communications NCC 2011. Indian Institute of Science, Bangalore
37. http://dimacs.rutgers.edu/Workshops/Video/abstracts.html
38. http://www.ee.columbia.edu/ln/dvmm/newHome.htm
39. http://www.merl.com/areas/VideoMining/
40. Hu W, Xie N, Li L, Zeng X, Maybank S (2011) A survey on visual content-based video indexing and retrieval. IEEE Trans Syst Man Cybern C: Appl Rev 1–23
41. Jiang F, Yuan J, Tsaftaris SA, Katsaggelos AK (2011) Anomalous video event detection using spatiotemporal context. Int J Comput Vis Image Underst 115:323–333
42. Jiang F, Yuan J, Tsaftaris SA, Katsaggelos AK (2011) Anomalous video event detection using spatiotemporal context. Comput Vis Image Underst 115:323–333
43. Jiang S, Tian Y, Huang Q, Huang T, Gao W (2009) Content-based video semantic analysis. Semantic mining technologies for multimedia databases. IGI Press
44. Kokkoras F, Jiang H, Vlahavas I, Elmagarmid AK, Houstis EN, Aref WG (2002) Smart VideoText: a video data model based on conceptual graphs. ACM Multimedia Syst J 8(4):328–338
45. Kotsiantis S, Kanellopoulos D, Pintelas P (2004) Multimedia mining. WSEAS Trans Syst 10(3):3263–3268
46. Kucuk D, Yazici A (2011) Exploiting information extraction techniques for automatic semantic video indexing with an application to Turkish news videos. Int J Knowl-Based Syst 24:844–857
47. Kea J, Zhana Y, Chenc X, Wanga M (2012) The retrieval of motion event by associations of temporal frequent pattern growth. Future Generation Comput Syst (in press)
48. Kiran Sree P (2008) Video data mining framework for information retrieval. In: Proceedings of NCKM-2008, Annamalai University, Tamilnadu, India
49. Kumar P, Roy S, Mittal A, Kumar P (2008) On-line data management framework for multimedia surveillance system. In: Proceedings of national conference on communications, 01–03 Feb 2008. IIT, Bombay
50. Lai C, Rafa T, Nelson DE (2006) Mining motion patterns using color motion map clustering. SIGKDD Explor 8(2):3–10
51. Lu S, Lyu MR, King I (2004) Video summarization by spatial-temporal graph optimization. Proc IEEE Int Symp Circuits Syst 2:197–201
52. Lin L, Shyu M-L (2010) Weighted association rule mining for video semantic detection. Int J Multimedia Data Eng Manag 1(1):37–54
53. Leavitt N (2002) Let's hear it for audio mining. Computer-Magazine
54. Lee J, Rajauria P, Shah SK (2007) A model-based conceptual clustering of moving objects in video surveillance. In: Proceedings of SPIE IS&T electronic imaging. SPIE, vol 6506, Jan 28–Feb 1, 2007, San Jose

55. Lin L, Ravitz G, Shyu M-L, Chen S-C (2007) Video semantic concept discovery using multimodal-based association classification. ICME07
56. Liu H-Y, He T (2009) Semantic event mining in soccer video based on multiple feature fusion. In: Proceedings of international conference on information technology and computer science, pp 297–301
57. Moxley E, Mei T, Manjunath BS (2010) Video annotation through search and graph reinforcement mining. IEEE Trans Multimedia 12(3):184–194
58. Maheshkumar HK, Palaniappan K, Sengupta S, Seetharaman G (2009) Semantic concept mining based on hierarchical event detection for soccer video indexing. J Multimedia 4(5):298–307
59. Ma YF, Lu L, Zhang, HJ, Li M (2002) A user attention model for video summarization. In: Proceedings of the tenth ACM international conference on multimedia, pp 533–542
60. Marsala C, Detyniecki M (2003) Fuzzy data mining for video. In: Proceedings of the international conference of the European society for fuzzy logic and technology—EUSFLAT'2003, pp 73–80
61. Matsuo Y, Amano M, Uehara K (2002) Mining video editing rules in video streams. In: Proceedings of the tenth ACM international conference on multimedia, pp 255–258
62. Mei T, Ma Y-F, Zhou H-Q, Ma W-Y, Zhang H-J (2005) Sports video mining with mosaic. In: Proceedings of 11th international multimedia modelling conference (MMM'05), Melbourne, Australia, pp 107–114
63. Mihajlovic V, Petkovic M (2001) Automatic annotation of Formula 1 races for content-based video retrieval. CTIT Tech Rep Series, TR-CTIT 01-41
64. Missaoui R, Palenichka RM (2005) Effective image and video mining: an overview of model-based approaches. In: Proceedings of 6th international workshop on multimedia data mining: mining integrated media and complex data, pp 43–52
65. Mitra S, Tinkuacharya (2003) Data mining multimedia, soft computing, and bioinformatics. Wiley, Hoboken
66. Mongy S, Bouali F, Djeraba C (2005) Analyzing user's behavior on a video database. In: Proceedings of the 6th ACM international workshop on multimedia data mining: mining integrated media and complex data (MDM/KDD 2005), pp 95–100
67. Moxley E, Mei T, Hua XS, Ma W-Y, Manjunath BS (2008) Automatic video annotation through search and mining. In: Proceedings of IEEE international conference on multimedia and expo (ICME), pp 685–688
68. Naphade MR, Huang TS (2001) A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Trans Multimedia 3(1):141–152
69. Nemrava J, Svátek V, Buitelaar P, Declerck T (2008) Text mining as support for semantic video indexing and analysis. In: Proceedings of the 2nd K-space PhD Jamboree workshop, Paris, France, July 25, 2008
70. Ngo C-W, Ma Y-F, Zhang H-J (2003) Automatic video summarization by graph modeling. In: Proceedings of the ninth IEEE international conference on computer vision (ICCV 2003), vol 2, pp 104–109
71. Nowozin S, Bakir GH, Tsuda K (2007) Discriminative subsequence mining for action classification. In: Proceedings of eleventh IEEE international conference on computer vision (ICCV2007), pp 1–8
72. Oh J, Bandi B (2002) Multimedia data mining framework for raw video sequences. In: Proceedings of the third international workshop on multimedia data mining (MDM/KDD'2002) in conjunction with the eight ACM SIGKDD international conference on knowledge discovery & data mining, pp 1–10
73. Oh J, Lee J, Hwang S (2005) Video data mining. Idea Group Inc
74. Oh J, Lee J, Kote S (2003) Real time video data mining for surveillance video streams. In: Proceedings of the seventh Pacific-Asia conference on knowledge discovery and data mining, pp 222–233
75. Oh JH, Wen Q, Hwang S, Lee J (2005) Video abstraction. Video data management and information retrieval, chap XIV. Idea Group Inc, IRM Press
76. Okada Y, Tada T, Fukuta K, Nagashima T (2010) Audio classification based on a closed itemset mining algorithm. Proc IEEE 2010:60–65
77. Pan J-Y, Faloutsos C (2001) VideoGraph: a new tool for video mining and classification. In: Proceedings of joint conference on digital libraries (JCDL'01), pp 116–117
78. Pan J-Y, Faloutsos C (2002) VideoCube: a novel tool for video mining and classification. In: Proceedings of the fifth international conference on Asian digital libraries (ICADL 2002), pp 194–205
79. Petkovic M, Jonker W (2001) Content-based retrieval of spatio-temporal video events. In: Proceedings of multimedia computing and information management track of IRMA international conference
80. Quack T, Ferrari V, Gool LV (2006) Video mining with frequent itemset configurations. In: Proceedings of international conference on image and video retrieval (CIVR (2006), Tempe, AZ, USA
81. Rasheed Z, Sheikh Y, Shah M (2005) On the use of computable features for film classification. IEEE Trans Circuit Sys Video Technol 15(1):52–64
82. Rasheed Z, Shah M (2003) Video categorization using semantics and semiotics, chap 7. Video mining. Kluwer, Dordrecht
83. Rui Y, Huang TD (2000) A unified framework for video summarization. Browsing and retrieval, image and video processing handbook, pp 705–715
84. Snoek CGM, Worring M (2005) Multimodal video indexing: a review of the state-of-the-art. Multimedia Tools Appl 25(1):5–35
85. Salway A (1999) Video annotation: the role of specialist text. PhD dissertation, Department of Computing, University of Surrey
86. Sanjeevkumar RJ, Praveenkumar K (2007) Multimedia data mining in digital libraries: standards and features. In: Proceedings of conference on recent advances in Information Science & Technology (READIT-2007), pp 54–60
87. Shirahama K, Ideno K, Uehara K (2005) Video data mining: mining semantic patterns with temporal constraints from movies. In: Proceeding of seventh IEEE symposium on multimedia, pp 598–604
88. Shirahama K, Iwamoto K, Uehera K (2004) Video data mining: rhythms in a movie. In: Proceedings of IEEE international conference on multimedia and expo, vol 2, pp 1463–1466
89. SivaSelvan B, Gopalan NP (2007) Efficient algorithms for video association mining. In: Proceedings of the 20th conference of the Canadian Society for computational studies of intelligence on advances in artificial intelligence, vol 4509. Springer, Berlin, pp 250–260
90. Sivic J, Zisserman A (2004) Video data mining using configurations of viewpoint invariant regions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 488–495
91. Su J-H, Huang Y-T, Tseng VS (2008) Efficient content-based video retrieval by mining temporal patterns. In: Proceedings of 9th international workshop on multimedia data mining associated with the ACM SIGKDD 2008, pp 36–42
92. Suresh V, Krishna Mohan C, Kumaraswamy R, Yegnanarayana B (2005) Combining multiple evidence for video classification. In: Proceedings of IEEE international conference on intelligent sensing and information processing (ICISIP-05), pp 187–192
93. Thuraisingham BM (2001) Managing and mining multimedia databases. CRC Press, Boca Raton

94. Tien M-C, Wang Y-T, Chou C-W, Hsieh K-Y, Chu W-T, Wu J-L (2008) Event detection in tennis matches based on video data mining. Proc ICME 2008:1477–1480
95. Tseng VS, Su J-H, Huang J-H, Chen C-J (2008) Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation. IEEE Trans Multimedia 10(2):260–268
96. Tien M-C, Wang Y-T, Chou C-W, Hsieh K-Y, Chu W-T, Wu J-L (2008) Event detection in tennis matches based on video data mining. In: Proceedings of ICME, pp 1477–1480
97. Turaga PK, Veeraraghavan A, Chellappa R (2007) From videos to verbs: mining videos for activities using a cascade of dynamical systems. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
98. Vibha L, Chetana Hegde P, Shenoy D, Venugopal KR, Patnaik LM (2008) Dynamic object detection, tracking and counting in video streams for multimedia mining. IAENG Int J Comput Sci 35:3
99. Vailaya A, Jain A, Zhang H (1996) Video clustering. Tech Rep. No. MSU-CPS-96-64, Michigan State University, USA
100. Vassiliadis B, Stefani A, Drossos L, Ioannou K (2005) Knowledge discovery in multimedia repositories: the role of metadata. In: Proceedings of 7th WSEAS international conference on mathematical methods and computational techniques in electrical engineering, pp 330–335
101. Wang F, Lu W, Liu J, Shah M, Xu D (2008) Automatic video annotation with adaptive number of key words. In: Proceeding of 19th international conference on pattern recognition (ICPR 2008), pp 1–4
102. Wang Y, Xing C, Zhou L (2006) Video semantic models: survey and evaluation. Int J Comput Sci Network Sec 6(2A):10–21
103. Weber J, Lefever S, Gancarski P (2010) Video object mining: issues and perspectives. In: Proceedings of IEEE fourth international conference on semantic computing, pp 85–91
104. Wu C, He Y, Zhao L, Zhong Y (2002) Motion feature extraction scheme for content-based video retrieval, storage and retrieval for media databases. Proc SPIE 4676:296–305
105. Xiong Z, Zhou XS, Tian Q, Rui Y, Huang TS (2006) Semantic retrieval of video. IEEE Signal Process Mag 23(2):18–27
106. Xie L, Chang S-F (2006) Pattern mining in visual concept streams. In: Proceedings of IEEE international conference on multimedia and expo (ICME06), Toronto, Canada
107. Yahiaoui I, Merialdo B, Huet B (2006) Automatic video summarization. Interactive video algorithms and technologies. Springer, Berlin
108. Yang X-F, Tian Q, Xue P (2007) Efficient short video repeat identification with application to news video structure analysis. IEEE Trans Multimedia 9(3):600–610
109. Zhang C, Chen W-B, Yang L, Chen X, Johnstone JK (2007) Automatic in vivo microscopy video mining for leukocytes. SIGKDD Explor 9(1):30–37
110. Zhou L, Shi Y, Feng J, Sears A (2005) Data mining for detecting errors in dictation speech recognition. Trans Speech Audio Process 13(5):681–689
111. Zhu X, Wu X, Elmagarmid A, Feng Z, Wu L (2005) Video data mining: semantic indexing and event detection from the association perspective. IEEE Trans Knowl Data Eng 17(5):1–14
112. Zang Q, Klette R (2003) Object classification and tracking in video surveillance. Communication and Information Technology Research Technical Report 128
113. Zhang Z (2002) Mining surveillance video for independent motion detection. In: Proceedings of IEEE international conference on data mining (ICDM), Maebashi City, Japan, Dec 2002
114. Zhao N, Chen SC, Shyu M-L (2006) Video database modeling and temporal pattern retrieval using hierarchical Markov model mediator. In: Proceedings of first IEEE international workshop on multimedia databases & data Management, pp 10–19
115. Zhu X, Aref W, Fan J, Catlina A, Elmagarmid A (2003) Medical video mining for efficient database indexing, management and access. In: Proceedings 19th international conference on data engineering, issue 5–8, pp 569–580
116. Zhu X, Fan J, Aref W, Elmagarmid A (2002) ClassMiner: mining medical video content structure and events towards efficient access and scalable skimming. In: Proceedings of ACM SIGMOD workshop on research issues in data mining and knowledge discovery, Madison, Wisconsin, USA, pp 9–16
117. Zhu X, Fan J, Hacid M-S, Elmagarmid A (2002) ClassMiner: mining medical video for scalable skimming and summarization. In: Proceedings of multimedia'02, Juan-les-Pins, France