ISSN (Online): 2230-7893
www.IJCEM.org

A Survey on Multimedia Data Mining

Shirish B. Patil[1],Saurabh Bonde[2], B.B.Meshram[3]

[1]Computer Department ,VJTI
Matunga (E),Mumbai
shirishpatil7@gmail.com

[2]Computer Department ,VJTI
Matunga (E),Mumbai
saurabh_bhonde@yahoo.in

[3]Head Of Computer Department,VJTI
Matunga (E),Mumbai
bbmeshram@vjti.org.in

Abstract

Multimedia Data Mining is the sub branch of Data Mining. Multimedia Data mining further classified according to the multimedia data type that are image, audio, video. Mostly, Multimedia data mining covers the mining of image data, mining of audio data and mining of video data. Image-driven data mining methods are described for image content classification, segmentation and attribution, where each pixel location of an image-under-analysis is the centre point of a pixel-block query that returns an estimated class label. Feature attribute estimates may also be mined when sufficient attribute strata exist in the data warehouse. Novel methods are presented for pixel-block mining, pattern similarity scoring, class label assignments, and attribute mining. Audio mining is a technique by which the content of an audio signal can be automatically analyzed used and searched. It is most commonly used in the field of automatic speech recognition, where the analysis tries to identify any speech within the audio. User has needed to retrieve meaningful information from their digital collections. To help users find and retrieve relevant video more effectively and to facilitate new and better ways of entertainment, advanced technologies must be developed for indexing, filtering, searching, and mining the vast amount of videos.

Keywords- *Data mining, Knowledge discovery, Image mining, Audio mining, Video mining, pattern matching, association rules.*

1. Introduction

Data Mining is the science of extracting useful information from large datasets or databases. It is technique of analyzing data. Data mining is the process of automating information discovery which improves decision making.

Multimedia Data mining is the sub-branch of Data Mining which deals with the multimedia data such as image, audio, video, etc. Subfield of data mining that deals with an extraction of implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia databases. Multimedia data mining is the part of multimedia technology. It covers areas such as media compression and storage, Delivering streaming media over networks with required quality of service, Media indexing, summarization, search and retrieval, Creating interactive multimedia systems for learning and creative art production, creating multimodal user interfaces. Multimedia includes a combination of text, audio, still images, animation, video, and interactivity content forms.

1.1. Representation of Multimedia Data

Representation of multimedia data means how the data will be available and in what form it is available**.** Image data contains black and white and color images. Audio data contains sound, speech and music. Video Data includes time aligned sequence of images.

1. Text

Plain text consists of alphanumeric characters. Optical character recognition (OCR) techniques are applied to convert analog text to digital text. The most common digital representation of characters is the American Standard Code for Information Interchange (ASCII).

The required storage space for a text document is equal to the number of characters. Structured text documents are becoming more and more popular. Such a document consists of titles, chapters, sections, paragraphs, and so on. A title may be presented to the user in a format different from a paragraph or a sentence.

ISSN (Online): 2230-7893
www.IJCEM.org

Standards like HTML and XML are used to encode structured information. There are some techniques, e.g., Huffman and arithmetic coding to compress text, but as storage requirements are not too high, the compression techniques are in general less important for text than for multimedia data.

2. Image

Digital images can be obtained by scanning (analog) photos and pictures using a scanner. The analog image is approximated by a rectangle of small dots. Another source of digital images is formed by the frames of a digitized or digital video. Images can be in gray-scale or in color. An image displayed on a screen consists of many small dots or picture elements (pixels). To describe the gray scale of a pixel we need say one byte of eight bits. For a color pixel we need three colors (e.g., Red, Green, and Blue) of one byte each. So, for a rectangular screen we can compute the amount of data required for the image using the formula

$$A = xyb$$

Where A is the number of bytes needed,

x is the number of pixels per horizontal line, y is the number of horizontal lines, and b is the number of bytes per pixel. Image compression is based on exploiting redundancy in images and properties of the human perception. It appears that pixels in a certain area are often similar; this is called *spatial redundancy*. Several compression techniques are available; among others transform coding, and fractal image coding.

3. Audio

Audio is caused by air pressure waves having a frequency and amplitude. When the frequency of the waves is between 20 to 20,000 Hertz a human hears a sound. Besides frequency, also the amplitude of a wave is important. Low amplitude causes the sound to be soft. How to digitize these pressure waveforms? The waveform can be digitized as, First, the air wave is transformed into an electrical signal (by a microphone). This signal is converted into discrete values by processes called *sampling* and *quantization*.

Sampling causes the continuous time axis to be divided into small, fixed intervals. The number of intervals per second is called the sampling rate. The determination of the amplitude of the audio signal at the beginning of a time interval is called quantization. So the continuous audio signal is approximated by a sequence of values. If the sampling rate is high enough and the quantization is precise enough the human ear will not notice any difference between the analog and digital audio signal. The process just described is called analog-to-digital conversion (ADC); the other way around is called digital-to-analog conversion (DAC).

4. Video

A digital video consists of a sequence of frames or images that have to be presented at a fixed rate. Digital videos can be obtained by digitizing analog videos or directly by digital cameras. Playing a video at a rate of 25 frames per second gives the user the illusion of a continuous view. It takes a huge amount of data to represent a video. In general the image compression techniques can also be applied to the frames of videos. The same principles as with images are used: reducing redundancies and exploiting human perception properties. Besides spatial redundancy we also have *temporal redundancy*. This means that neighboring frames are normally similar. This redundancy can be removed by applying the motion estimation and compensation where each image is divided into fixed-size blocks. For each block in the current image the most similar block in the previous image is determined and the *pixel difference* is computed. Together with the *displacement* between the two blocks, this difference is stored and if needed transmitted.

2. Related work

2.1 Multimedia Data mining Architecture

Multimedia Data mining architecture follows the seven steps that are: Domain understanding stage, Data selection, Cleaning and Preprocessing, Pattern-Discovery, Interpretation, Reporting and putting

1. Domain Understanding Stage

Domain understand stage requires learning how the results of data mining will be used so as to gather all relevant prior knowledge often leads to discovery of irrelevant or meaningless patterns .e.g. Cricket, it is very important to have a good knowledge and understanding of the game to detect interesting strokes used by batsman

2. Data Selection

The data selection stage requires the user to target a database or select a subset of fields or data records to be used for data mining. The proper domain understanding at this stage helps in the identification of useful data. This is the most time consuming stage of the entire data mining process for business applications.

For multimedia data mining this stage is not an issues because data are not in relational form and there are no subsets of fields to choose form.

3. Cleaning and Preprocessing

In preprocessing stage involves integrating data from different sources and making choices about representing or coding certain data fields that serve as inputs to the pattern discovery stage.

ISSN (Online): 2230-7893
www.IJCEM.org

The preprocessing stage is of considerable importance in multimedia data mining, given the unstructured nature of multimedia data.

4. Pattern-Discovery

The pattern discovery stage is the heart of entire data mining process. It is the stage where the hidden patterns and trends in the data are actually uncovered. There are several approaches to the pattern discovery stage these include association, clustering, regression, time series analysis and visualization. Each of these approaches can be implemented through one of several coating methodologies, such as statistical data analysis, machine learning, neural networks and pattern recognition.

5. Interpretation

The interpretation stage of the data mining process is used to evaluate the quality of discovery and its value to determine whether previous stage should be revised or not. Proper domain understanding is crucial at this stage to put a value on discovered patterns.

6. Reporting and putting

The final stage of the data mining process consists of reporting and putting to use the discovered knowledge to generate new actions or products and services or marketing strategies as the case may be.

Architecture captures all above stages of data mining in the context of multimedia data. Figure shows the whole architecture of multimedia data mining. The broken arrows on the left indicate that the process is iterative.
The arrows emanating from the domain knowledge block on the right indicate domain knowledge guides in certain stages of the mining process.

The spatiotempororal segmentation step is necessary because of the unstructured nature of the data. This step breaks multimedia data into parts that can be characterizes in terms of certain attributes or features.
In conjunction with the feature extraction step, this step serves the function similar to that of preprocessing step in typical data mining process. In image data mining the spatiotemporal step simply involves image segmentation. Both region and edge-based image segmentation methods have been used at this stage in different applications.

3.Discussion

Multimedia Data Mining is classified according to the multimedia data type.
Brief classification of Multimedia data mining is:

3.1 Image Data Mining
3.2 Audio Data Mining
3.3 Video Data Mining

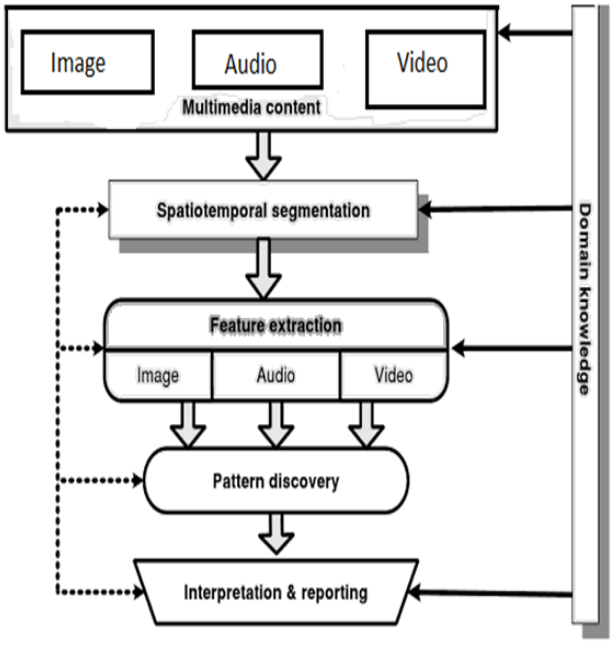


Fig1. Multimedia Data Mining architecture

3.1 Image data mining

Image-driven data mining methods are described for image content classification, segmentation and attribution, where each pixel location of an image-under-analysis is the centre point of a pixel-block query that returns an estimated class label. Feature attribute estimates may also be mined when sufficient attribute strata exist in the data warehouse. Novel methods are presented for pixel-block mining, pattern similarity scoring, class label assignments, and attribute mining.

1.Pattern Matching With σ-Trees

A progression of successive approximation pattern recognition decisions are obtained by searching and comparing the bit-plane indexes of variable depth XDRs $\mathbf{i}p = (i1,i2, \ldots . \ldots ,ip)$ of archived referential knowledge base content and XDR of the content of an image being classified.

ISSN (Online): 2230-7893
www.IJCEM.org

The set of matched bit-plane indexes **i***p* are subsequently used for data mining extraction of feature class member estimates, and sometimes, feature attribute value estimates if the required attribute strata are available in the data warehouse.

Figure shows Four-element pixel-block-aggregate source-optimized *σ-tree templates* and classifier structure. *σ-Tree Classifier Shown at the* top left /top-right *works as an Indexing Engine.* An *σ-tree design method* is applied to generate a three-stage *σ-tree with two templates* at each stage (shown on the bottom left in Fig.).These six templates are stored and used "on-the-fly" to form the direct sum *σ-tree structure shown on the bottom right* .The pixel values of the residual pattern templates, as shown on the lower-left quadrant in Fig., may have positive or negative values.
The four training pixel blocks (shown again on the top right in Fig.) are input into the *σ-tree classifier to* generate XDRs.

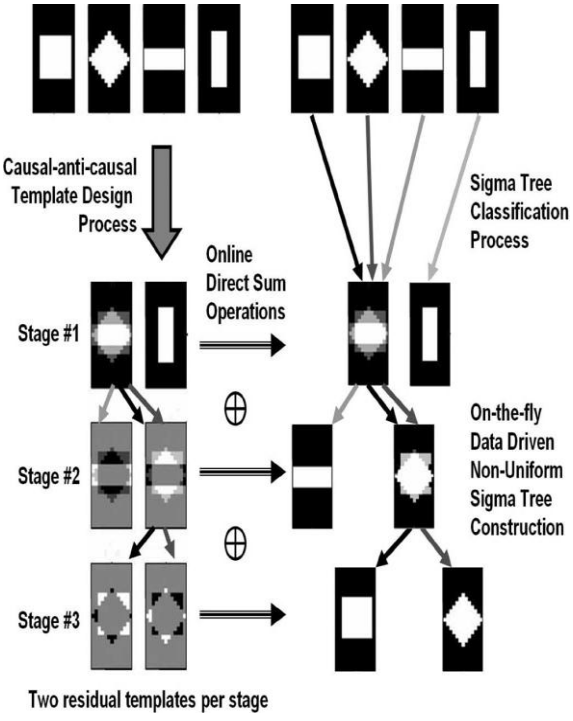


Fig2:-Four-element pixel-block-aggregate source-optimized *σ-tree templates* and classifier structure.

The arrows that connect the templates in Fig. show the sequential search path for each pixel-block query. First warehouse member =t(1) on the left is encoded through a left–right–left *σ-tree path sequence, giving an XDR of σ3(t) = i3(1) = (010).*

Second warehouse member from the left t(2) is encoded through a left–right–right path sequence *giving an XDR of σ3(t(2)) = i3(2) = (011).*

Third member from the left is encoded through a left–left–stop sequence *σ3(t(3)) = i2(3)* =(00*x*)*, with x indicating a truncated encoding* .

The aggregate member on the right is encoded through a right–stop sequence *σ3(t(4)) = i1(4)* =(1*xx*)*.* by this example is that the storage and searching of *σ-tree stage templates Corresponds to exponential growth in* the corresponding options of *σ-tree decision outcomes.* In this example, three binary stages result in 23 terminating leaf nodes.

2. Data mining with σ TREES

First, assume a *feature locator table* **L***a* stored in a database. Each row of the feature locator table describes the location of an archived feature exemplar by specifying a source image **I**source, the pixel-offset location (*X, Y*) of a pixel block containing the feature exemplar, a descriptive feature label **F**, and any attribute labels or attribute values **A** that may be known about the feature example. The locator data-tuples can thus be described as **L***a* = [*k,* **I**source,*X, Y*,**F**,**A**], where *k* is the primary key of the database table **L***a*, and *A* is used to differentiate feature locator tables.

A pixel-block extraction tool combined with an extraction query *Qu* into **L***a* is used to form a *snippet set* **S***Qu* of pixel block feature exemplars to be used for *σ*-tree training.

3. Image-Driven Mining Systems
Image-driven mining of each pixel location of an image under-analysis returns a class label **F**MAP; and when additional strata data **A** about feature attributes are available in the warehouse, estimated attribute values can also be returned by the data mine search.

Nearly all of the required tools are in place to describe the basic flow of image-driven mining for feature extraction and autonomous attribution. However, first, a method of extracting blocks from new images to feed to the pixel-block query search engine is needed—this is easily done with a sliding window algorithm.
Given a new image **I**name, a sliding window moves through the image and extracts pixel blocks **t**(*x, y*) of size (*bw, bh, bd*) from the sliding window positions (*x, y*). The extracted pixel groups (blocks or cubes) are called "snippets."

The sliding window is moved through every possible location in the input image. Hence, all possible snippets are extracted from the image (except for image edge boundary effects) and the entire query image can be processed for IDDM. The stride of the sliding window can be adjusted to return computation efficiency for a trade-off of a less dense query grid.

The following describe the main preparatory and exploitive steps of a SOLDIER IDDM process.

1) Build a warehouse by creating knowledge base feature locator tables **L***a*.

2) Specify snippet pixel-block size (*bw*, *bh*, *bd*) and query against the locator table **L***a* to form a pixel-block aggregate **S***Qu*.

3) Build σ-tree $T\sigma$ from **S***Qu* and form the SOLDIER Index Table $T\sigma$(**S***Qu*) for the archived feature aggregate **S***Qu*.

4) Build SOLDIER maximum posterior index hash tables *Hp*[**i***p*,**F**MAP(*p*)].

5) Apply sliding window on new image **I**query to extract pixel-block queries for IDDM.

6) Perform the image-driven data mine by using $T\sigma$ to find $\hat{\mathbf{t}}$ (x,y)(**i***p*) : $p = 1, \ldots, P$ for each **t**(x, y) $\in$ **I**query, and then use the resultant **i***p* from the image-under-analysis to data mine against the warehouse **i***p* $\in T\sigma$ from **S***Qu* columns of all of the stage hash tables. If commanded, in a user-interactive mode, data mine for attribute information with mouse clicks on pixel-block query returns to explore attribute strata **A** via structured query language (SQL) formatted queries.

3.2 Audio Data Mining

Audio mining is a technique by which the content of an audio signal can be automatically analyzed and searched. It is most commonly used in the field of automatic speech recognition, where the analysis tries to identify any speech within the audio.

This information may either be used immediately in pre-defined searches for keywords or phrases or the output of the speech recognizer may be stored in an index file. Audio mining systems used in the field of speech recognition are often divided into two groups: those that use Large Vocabulary Continuous Speech Recognizers (LVCSR) and those that use phonetic recognition.

1. Audio mining approaches

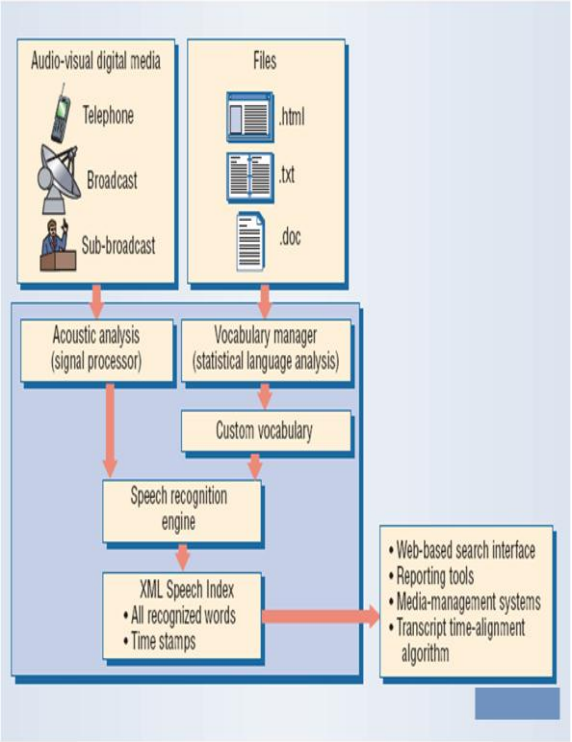
There are two main approaches to audio mining such as Text-based indexing and Phoneme-based indexing.

Text-based indexing, also known as large – vocabulary continuous speech recognition. It converts speech to text and then identifies words in a dictionary that can contain several hundred thousand entries. If a word or name is not in the dictionary, the LVCSR system will choose the most similar word it can find.

Phoneme based indexing doesn't convert speech to text but instead works only with sounds. The system first 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.

2. Speech detection System

The Audio Mining Development System works with audio from various sources. The system analyzes sounds to generate sound strings that are then identified as words by the speech recognition engine, which works with its own dictionary. Speech recognition engine converts spoken words to text. Material input into the vocabulary manager automatically updates the dictionary. The product's XML Speech Index uses XML's cross-platform capabilities to create files that work with various search engines, servers, and content management systems.



In view of the fact that an audio element can consist of speech, music, various audio effects, or any combination of these, high discriminative capabilities are required for the used feature set. Speech Engine is an accurate, standards-based speech recognizer that supports multiple languages and can perform speech recognition on audio data from any audio source.

3.3 Video Data Mining

Organizations with large digital assets have a need to retrieve meaningful information from their digital collections. To help users find and retrieve relevant video more effectively and to facilitate new and better ways of entertainment, advanced technologies must be developed for indexing, filtering, searching, and mining the vast amount of videos. A simple framework is to partition continuous video frames into discrete physical shots and extract low-level features from video shots to support activities like searching, indexing or retrieval. The purpose of video data mining is to discover and describe interesting patterns in data. The task becomes especially tricky when the data consist of video sequences (which may also have audio content), because of the need to analyze enormous volumes of multidimensional data. The New Video Data Mining Technology measures emotional response of shoppers to products at the Shelf. This new technology uses automatic facial expression as well as behavior analysis to rate the emotional response and interest of shoppers for a particular product.

1 Video mining approaches

Three kinds of video mining approaches**:**

1. **Special pattern detection** which detects special patterns that have been modeled in advance, and these patterns are usually characterized as video events (e.g., dialog, or presentation).
2. **Video clustering and classification** which clusters and classifies video units into different categories. For example, in video clips are grouped into different topic groups, where the topic information is extracted from the transcripts of the video.
3. **Video association mining**, where associations from video units are used to explore video knowledge

An intuitive solution for video mining is to apply existing data mining techniques to video data directly. Nevertheless, as we can see from the three types of video mining techniques above, except, which have integrated traditional sequential association mining techniques, most others provided their own mining algorithms. The reason is that almost all existing data mining approaches deal with various databases (like transaction data sets) in which the relationship between data items is explicitly given. The greatest distinction between video and image databases is that the relationship between any two of their items cannot be explicitly figured out. Although we may now retrieve video frames (and even physical shots) with satisfactory results, acquiring relationships among video frames (or shots) is still an open problem. This inherent complexity has suggested that mining knowledge from multimedia materials is even harder than from general databases.

2. KNOWLEDGE-BASED VIDEO INDEXING

A knowledge-based video indexing framework is used for video database management and access. There are two widely accepted approaches for accessing video in databases: shot-based and object-based.

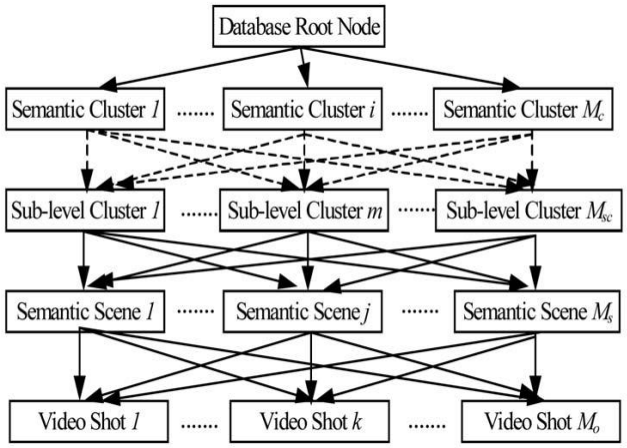


Fig3.(a) The proposed hierarchical video database model.

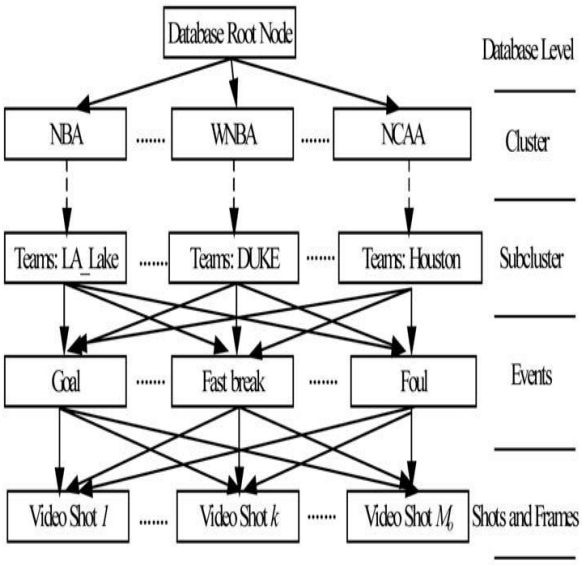


Fig3.(b) Knowledge-based basketball video database management.

In comparison with traditional video database systems that use low-level similarities among shots to construct indices, a semantic video database management framework has been proposed in Fig3. (a), where video semantic units (scenes or story units) are used to construct database indices. However, this scheme works on videos with content structure, e.g., movies and news, where video scenes are used to convey scenarios and content evolution. For many other videos, such as sports videos, there are no such story units. Instead, they contain various interesting events, e.g., a goal or a fast break, which could be taken as highlights and important semantics. Accordingly, by integrating the existing framework in Fig3. (a). we propose a knowledge-based video indexing framework for basketball videos, as shown in Fig3.(b). To support efficient video indexing, we need to address the following three key problems before we can actually adopt the framework in Fig3. (b):

1) How many levels should be included in the model?
2) Which kinds of decision rules should be used at each node?
3) Do these nodes make sense to human beings?

We solve the first and third problems by deriving knowledge from domain experts (or from extensive observations) and from the video concept hierarchy. For basketball videos, we first classify them into a two-level hierarchy. The first level is the host association of the games, e.g., NBA, NCAA, and CBA, and the second level consists of teams of each association,. Then, we integrate the structure of video content to construct lower-level indices. As we have stated above, extensive observations and existing research efforts suggest that there are many interesting events in sports videos that can be used as highlights. For basketball videos, the events that likely attract most viewers' interests are goals, fast breaks, and free throws, etc. We can therefore use these events as nodes at the third level of our indexing structure. At the lowest level, we use the video shots as index nodes, as shown in Fig., where each shot may have more than one parent node because some shots contain Several events association-based video indexing various features are outlined below:

1. A video association mining algorithm to discover video knowledge. It also explores a new research area in video mining, where existing video processing techniques and data mining algorithms are seamlessly integrated to explore video content.

2. An association-based video event detection scheme to detect various sports events for database indexing. In comparison with other video event detection Techniques, e.g., special pattern detection, the Hidden Markov Models, and classification rules, the association-based technique do not need to define event models in advance. Instead, the association mining will help us explore models (associated patterns) from video.

3. A knowledge-based sports video management framework to support effective video access. The inherent hierarchical video classification and indexing structure can support a wide range of granularity levels. The organization of visual summaries is also inherently supported. Hence, a naive user can browse only a portion of highlights (events) to get a concise summary.

4. Conclusions

Image data mining, Audio data mining and Video data mining are the sub branches of the multimedia data mining, classified according to the multimedia data type used. Image-driven data mining methods are described for image content segmentation, classification, and attribution, where each pixel location of an image-under-analysis is the centre point of a pixel-block query that returns an estimated class label. IDDM with trees has been demonstrated variety of image types and feature sets. In view of the fact that an audio element can consist of speech, music, various audio effects, or any combination of these, high discriminative capabilities are required for the used feature set. Speech Engine is an accurate, standards-based speech recognizer that supports multiple languages and can perform speech recognition on audio data from any audio source. We have used video associations to construct a knowledge based video indexing structure to support efficient video database management and access. In video data mining some video transformations are required before starting the actual mining of video data.

References

[1] IEEE paper on Data Mining: an Overview from a Database Perspective by Ming-Syan Chen,Jiawei Han,Philip S.Yu.

[2] IEEE journal paper on Image Driven Data mining for image Content Segmentation,Classification and attribution by Christopher F. Barnes.

[3] IEEE paper on Mining Multilevel image Semantics via Hierarchical Classification by Jianping Fan,Yuli Gao,Ramesh Jain,Hangzai Luo and Ramesh Jain

[4] IEEE paper on Video Semantic Events/Concept Detection Using a Subspace-Based Multimedia Data mining FrameWork by Ling Shyu,Min Chen,shu-Ching Chen

[5] IEEE paper on Video Semantic Events/Concept Detection Using a Subspace-Based Multimedia Data mining FrameWork by Ling Shyu,Min Chen,shu-Ching Chen.

[6] IEEE paper on Audio Keywords Discovery for Text-Like AudioContent Analysis and Retrieval by Lie Lu,Alan Hanjalic

[7] IEEE paper on Multimedia Data Mining and Its Implications for Query Processing by William I. Grosky and Yi Tao.