

## Knowledge Discovery using Various Multimedia Data Mining Technique

Miss. Ankita R. Makode  
Student, Computer Science and Engineering,  
P.R. Patil COET, Amravati,  
Maharashtra, India.

Prof. Arvind S. Kapse  
Assistant Professor, Computer Science and Engineering,  
P.R. Patil COET, Amravati,  
Maharashtra, India.

**Abstract:-** Knowledge discovery in databases (KDD) is the process of discovering positive information from a gathering of data. This generally used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior information on data sets and interpreting perfect solutions from the observed results. Knowledge Discovery in Databases is the process of finding knowledge in huge amount of data where data mining is the core of this process. Data mining can be used to understandable meaningful patterns from huge databases and these patterns may be transformed into knowledge.

Multimedia data mining can be defined as the process of finding motivating patterns from media data such as audio mining, video mining, image mining and text mining that are not generally available by basic queries and associated results. It is the mining of knowledge and high level multimedia information from large multimedia database system. Multimedia data mining refers to sample discovery, rule extraction and knowledge gaining from multimedia database.

In this paper, An Overview On various Multimedia Data technique is given and the main focus is given to the video Data Mining.

**Keywords:** data mining ,multimedia Data Mining ,text mining ,image mining ,audio mining ,video mining , clustering .

\*\*\*\*\*

### 1. Introduction

Multimedia data mining can be better understood by its purpose and scope. According to MPEG-7 Standard the kinds of data belong to the multimedia data [1] are of four types audio data, which includes sounds, speech, and music; image data, video data, which include time-aligned sequences of images; and electronic or digital ink, which is sequences of time aligned 2D or 3D coordinates of a stylus, a light pen, data glove sensors, graphical, temporal, relational and categorical data or a similar device are stored in a multimedia database and used to develop a multimedia system [1]

Definition1. — Multimedia data mining is the discovery and study, by automatic or semi-automatic means, of large quantities of data in order to discover important patterns and rules.

Definition2. —Multimedia data mining is a subfield of data mining that deals with an extraction of implicit knowledge, multimedia data connection or previous patterns not explicitly stored in multimedia database.

#### 1.1 WHAT IS MULTIMEDIA DATA MINING

Multimedia Data refers to data such as text, numeric, images, videos, audios, graphical, sound, temporal, files, animation, relational and categorical data. Multimedia data mining refers to sample discovery, rule extraction and knowledge acquisition from multimedia database [2]. The multimedia is classified in to two categories:(i) static media such as text, graphics, and images and (ii) dynamic media such as animation, musics, audio, speech, and video.

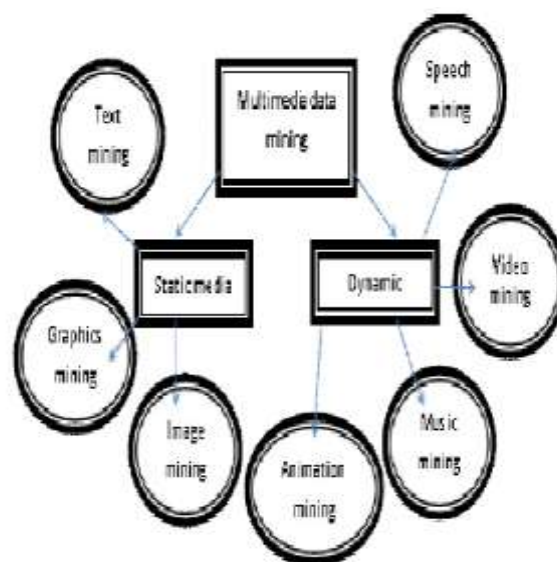


Figure 1.1: Multimedia data mining[3]

Figure 1.1 illustrates multimedia data mining, in particular, various aspects of multimedia data mining[2] Multimedia data mining approach for automated event study of suspicious movements Many government Data mining tool organizations, multi-nationals companies, shopping malls, banks and etc. are using surveillance system to monitor movements of employees, visitors, machines etc..

### 2. LITERATURE SURVEY-

Ja-Hwung Su [4] et al. The proposed an innovative method to achieve high quality of content-based video retrieval by discovering the temporal patterns in the video contents. On the basis of the discovered temporal patterns, THE efficient indexing technique and an effective sequence

matching technique were integrated to reduce the computation cost and to raise the retrieval correctness, respectively.

Sanjeevkumar R, [5] et al. The data mining techniques are useful while convert the multimedia files in the libraries. The digital library retrieves, collects, stores and preserves digital data. For this purpose, there is a need to transform different formats of information such as text, images, video, audio, etc. a data mining techniques are popular during the conversion of multimedia files in the libraries.

S. Kotsiantis, D. Kanellopoulos, P.Pintelas [6] describes that, the compare to data mining, multimedia data mining covers higher complexity resulting from: i) The large volume of data, ii) The inconsistency and heterogeneity of the multimedia data and iii) The multimedia content.

A.Hema,E.Annasaro[7] The explain All the views and ideas of all authors in field of the multimedia data mining. The need of image mining is mainly focused. Image mining have the great importance in the geological field, biological field, and pharmacy field. sample matching technique plays a vital role in Image mining. The process of extraction of useful knowledge hidden inside the image can be retrieve by sample matching technique also.

Pravin M. Kamde, Dr. Siddu P. Algur [8] The diagrammatical representation of mining multimedia database, text mining, image mining ,video mining, multimedia mining process are explain. Classification model, Clustering model and Association rule are some technique use for multimedia mining.

Kale D.V., Su [9] et al. the proposed a framework for surveillance videos of stationery places. They implemented an algorithm to group incoming video stream into meaningful pieces called segments. Further, they extracted a feature of segment which is used to characterize the segments. Motion of a segment is extracted using a two dimensional matrix which is constructed using accumulated pixel differences among all the frames in segment. Video segments are then clustered using K-means algorithm and finally found abnormality in the segments of video.

Chary [10]et al. the proposed evaluation of image retrieval methods. The retrieval images within a large image collection based on color projections and different mathematical approaches are introduced and applied for the retrieval images. Images are sub grouped using threshold values; they are considered R, G, B color combinations for retrieval images implemented. The results show that they are obtained efficient results compared to existing methods.

## 2. Hierarchical Structure of multimedia data mining

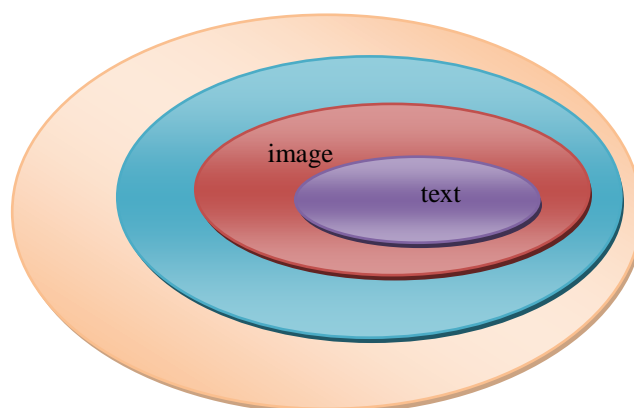


Figure 3.1 Hierarchical Structure Multimedia Data Mining

### 3.1. Multimedia data mining types

#### A. TEXT MINING:

Text Mining is a type of mining where data is extracted only in text format from the data bases. Text mining can be said as an extension Data mining. Here the data can be retrieved specifying the attributes or key words. [11]

#### B. IMAGE MINING:

Image mining systems that can be automatically extracted from image data are increasingly in demand. The fundamental challenge in the image mining is to determine how the low level, pixel demonstration contained in a raw image or image sequence can be processed to identify high-level spatial objects and correlation [11].

Image mining is the concept used to detect patterns and extract data from images stored in the large data bases.[11].

#### C. AUDIO MINING:

Audio is a continuous media type as video techniques used in audio are similar to the video data extraction. Audio can be in the form of radio, speech, etc.. To mine in audio data, first it has to be Converted into text using speech transcription techniques. Audio data can also be mined directly by using audio information techniques and then mining selects the audio data.

Audio mining is very simple in designing when compared to video mining. [3]Audio mining is a speaker-independent, speech recognition technique that Can Be used to search or recognize audio or video files. for occurrences of spoken words or Phonemes That are spoken speech recognition search engine identifies words or phonemes that are spoken within the file and generates a searchable index that includes a time

stamp for each important word or phoneme and its locations within the file

**D. Video Mining:**

Video contains a number of kinds of multimedia data for example text, image, metadata, visual and audio. It is regularly used in many major potential applications like security and surveillance, entertainment, medicine, culture programs and sports.[12] The Aim of video data mining is to discover and describe patterns from the large amount of video data as it is one of the core problem areas of data-mining research community. Compared to the mining of the additional types of data, video data mining is still in its immaturity [12].

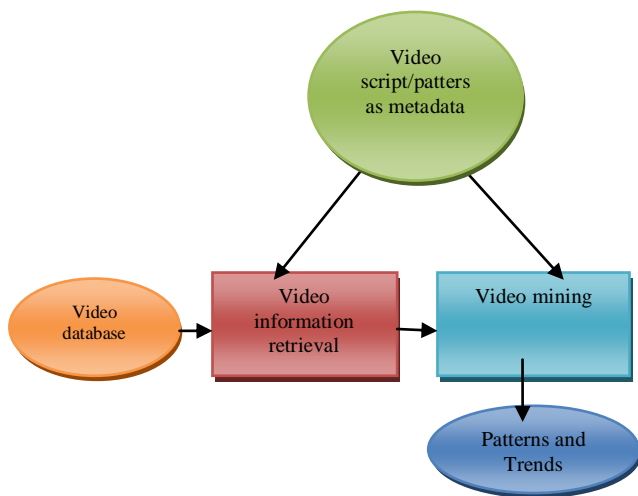


Figure 4. Video Mining[12]

**4. Table of multimedia data mining**

TYPE	FORMATS	TECHNIQUES AVAILABLE
IMAGE	GRAPHICS , ANIMATION	ASSOCIATION RULE, CLUSTERING
TEXT	KEYWORDS, PATTERNS	CLUSTERING
AUDIO	SPEECH , MUSIC	CLUSTERING
VIDEO	GAMMING	CLUSTERING

**5. multimedia data mining process**

The process of applying multimedia mining in order to retrieve various types of data is represented in Fig 5. Data collection is the first and foremost point of a learning system, as the excellence of raw data is the factor which determines the overall achievable performance. The main aim of data pre-processing is to determine. [1]

Given a training set, a knowledge model has to be chosen to learn from it and make multimedia mining model more iterative. Higher complexity found on compared data mining with multimedia mining: a) the large quantity of data, b) the

variability and heterogeneity of the multimedia data (e.g. diversity of sensors, time or conditions)[3]

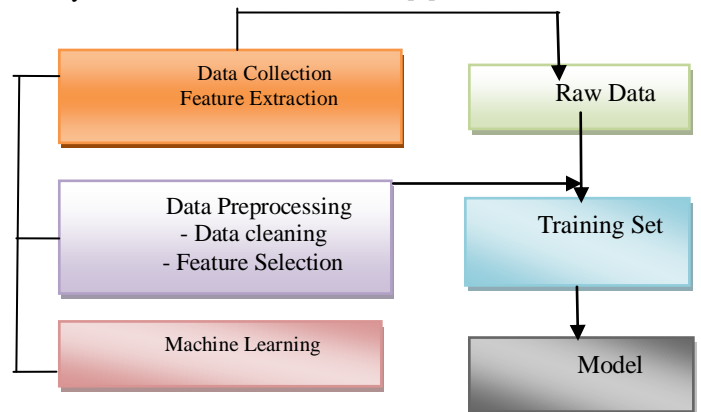


Figure 5 . multimedia data mining process[1]

present architecture of applying multimedia mining in different multimedia types Data collection is the starting point of a learning system, as the feature of raw data determines the overall achievable performance. Learning can be straightforward, if informative feature scan be identified at pre-processing stage. Detailed procedure depend highly on the nature of raw data and problem’s field. In some cases, prior information can be really valuable.[3]

**5.1. MULTIMEDIA DATA MINING TASKS [3, 1]**

**a. Multimedia Data Cube:**

model for multidimensional analysis of multimedia data; it is difficult to implement a data cube efficiently given the huge number of dimension. We may like to model color, course, texture, keywords, and so on. Many of the various dimensions in multimedia data cubes are of single, valued. e.g. one image may be matching to a set of keywords.

**b. Feature extraction:**

Multimedia features are extracted from media sequences or collections converting them into numerical or representational form. Good features shall be able to capture the perceptual saliency, distinguish contents mantic, as well as being computationally and represent atonally economical.

**c. Data Pre-processing:**

collecting data from similar sources and making choices about representing or codingcerta in data field is the task of this stage. It serves as input to the sample detection stage. Because certain fields may contain data at levels of details which are not considered suitable for the sample discovery stage representation other choices are needed.

**d. Discovering Patterns:**

The sample discovery stage is the heart of the entire data mining process. Several approaches of the pattern discovery

stage includes association, classification, clustering, regression, time-series study and visualization.

**e. Interpretations:**

To evaluate the feature of discovery and its value to determine whether previous stage should be revisited or not this stage of data mining process is used.

**f. Reporting and using discovered knowledge:**

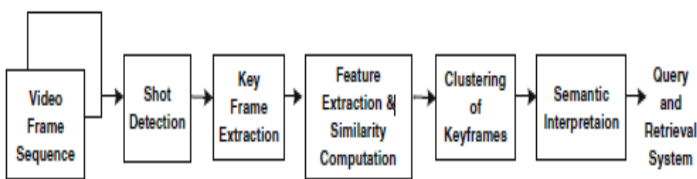
This Last stage is reporting and putting to use the discovered knowledge to generate actions or products and services or marketing strategies as the case may be. [1].

**5.2. TECHNIQUE FOR MULTIMEDIA DATA MINING:-**

**A. CLUSTERING:[3]**

Clustering is a useful technique for the discovery of some information from a data set it maps a data item into one of a number of clusters, where clusters are natural s collection for data items based on similarity metrics or probability density models( mitra & acharya ,2003).[13]

Cluster analysis divides the data object into several groups or clusters. Cluster analysis combines all objects based on there are mainly collections. Clustering algorithms can be divided into a number of method they are hierarchical method, density-based method, grid-based methods, and model based methods, k-means algorithm and graph based model . In multimedia mining, clustering technique can be applied to group related images, objects, sounds, videos and texts. There are two main techniques of audio mining. One uses large vocabulary continue speech recognition (LVCSR), and the other uses phonetic recognition to carry out phonetic audio mining.



**Figure 5.2 Video clustering[13]**

**6. CONCLUSION.**

In this paper, we addressed data mining for multimedia data such as text, image, video and audio. In particular, we have reviewed and analysed the multimedia data mining process with different tasks. This paper also described the clustering models using video for multimedia mining.

**REFERENCES:**

[1] Sarla More, and Durgesh Kumar Mishra ( march 2012)“Multimedia Data Mining: A Survey” Pratibha: International Journal Of Science, Spirituality, Business And Technology (IJSSBT) , ISSN (Print) 2277—7261, Vol. 1, No.1, pp. 49-55.

[2] Manjunath T. N, Ravindra S Hegadi , Ravikumar G K (November 2010)“ A Survey on Multimedia Data Mining and Its Relevance Today”,[ IJCSNS] International Journal of Computer Science and Network Security, VOL.10 No.11, PP-165-170.

[3] Manjunath R, S. Balaji , “Review and Analysis of Multimedia Data Mining Tasks and Models ” International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2 , pp 124-130, Special Issue 2, May 2014

[4] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh and Vincent Tseng (2010.), “Effective Content-Based Video Retrieval Using Pattern-Indexing and Matching Techniques”, Expert Systems With Applications, Vol. 37, No. 7, pp. 5068–5085,

[5] Sanjeevkumar R. Jadhav, and Praveenkumar Kumbargoudar (2007), “Multimedia Data Mining in Digital Libraries: Standards and Features” in Proc. READIT-, p. 54.

[6] Raymond J. Mooney and RazvanBunescu. “Mining Knowledge from Text Using Information Extraction”,SIGKDD Explorations. Volume 7, Issue 1.

[7] A.Hema, E.Annasaro, ( February 2013), “A SURVEY IN NEED OF IMAGE MINING TECHNIQUES”, International Journal of Advanced Research in Computer and Communication Engineering (IJARCC) ISSN (Print) : 2319-5940 ISSN (Online) : 2278-1021, Vol. 2, Issue 2.

[8] PravinM. Kamde, Dr.Siddu.P.Algur,( August 2011) “A SURVEY ON WEB MULTIMEDIA MINING”, the International Journal of Multimedia & Its Applications (IJMA) Vol.3,No.3.

[9] Kale D.V., Momin B.F.( May, June 2013) “Video Data Mining Framework for Surveillance Video”, International Journal of Advanced Trends in Computer Science and Engineering, Volume 2, No.3.

[10] R.Venkata Ramana Chary, Dr.D.Rajya Lakshmi and Dr. K.V.N, Sunitha,(March 2012.) “feature extraction methods for color image similarity”Advanced Computing :A International Journal (ACIJ ),Vol.3, No.2,

[11] Manda Jaya Sindhu, Y. Madhavi Latha, V. Samson Deva Kumar, Suresh Angadi(March 2013) “Multimedia Retrieval Using Web Mining” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, pp.106-108 .

[12] Priyanka A.Wankhade, Prof. Avinash P. Wadhe , ( April-May 2014 )“Survey on Analysis of Various Techniques for Multimedia Data Mining” International Journal Of Engineering Research And General Science (ISSN) 2091-2730 Volume 2, Issue 3, pp.137-142.

[13] V. Vijayakumar • R. Nedunchezian,( 2012)“ A study on video data mining”, Int J Multimed Info Retr ,1:153–172 DOI 10.1007/s13735-012-0016-2.