ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

A Study on Video Mining Using Different Tools

Dhiraj Paryani, Hariba Dhere, Sumit Lalwani, Nikhil Talreja, Dimple Bohra

Student, Dept. of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Student, Dept. of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Student, Dept. of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Student, Dept. of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Assistant Professor, Dept. of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

ABSTRACT: To create a system that mines the video data. In the recent years, the world had shown interest in video mining, which deals with the extraction of semantic info, implicit patterns and knowledge from video data to enhance the smart level of video applications. To create a system that understands the video and images gives precise description and provides a suitable title. Proposed system divide the video in set of images and audio. In order to avoid the same frames system uses an algorithm, which combines background subtraction with symmetrical differencing to get distinct frames from entire video[3].Then these frames(images) are clustered it using suitable clustering algorithm like K-means[1] or Fuzzy c-means. From these clusters, system obtain the different objects which are present in the frame .Then proposed system mines the video according to information that have been gathered from clustering and audio information using different data mining tools which are available.

I. INTRODUCTION

Video mining can be defined as a method to discover the knowledge, patterns, events of interest and structures from the video. The motivation for this comes from the tremendous success of data mining algorithms in discovering non-obvious patterns from data, which are not previously known. Video surfing which is supported by almost all systems, requires an update and modification because there is a chunk of data that is corrupted in such systems. When one search a video, he/she often get unwanted videos which are not related to search condition. This is because these system uses the title of the video to identify the video and the searching techniques are also based upon that keywords of title. So, by doing video mining basis on the content system will decide what video is about and system will provide the correct title for the video and by doing this system can achieve great efficiency in video surfing.

II. BACKGROUND

1. Video is nothing but images changing at very fast rate and audio plays according to timing. System extract the set of images and audio from video. Process take images from video according Fps. Fps means frames per second. So, Proposed system have to decide how many frames or images are taken out from the 1 second of video. Then that frames or images are processed individually. In order to avoid the same frames System uses an algorithm, which combines background subtraction with symmetrical differencing to get distinct frames from entire video[3].This algorithm takes two consecutive images and compare the two images according to intensity and checks whether the two images are similar or not. Hence at the end of this, it stores the set of images which are different or not similar to each other. Then, each image is passed through clustering algorithm. Clustering means grouping the image pixels into number of clusters which are related to each other based upon certain criteria. Clustering is a classification technique. Describing each pixel of an image by vector of N measurements. A similarity of the measurement vectors and therefore clustering in the N-dimensional measurement space implies similarity of the corresponding pixel groups or pixels. This means elements within a same cluster or group have high similarity while they differ from the elements in a different group. Therefore after clustering process, elements have high intra cluster property and low inter cluster property.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

2. There are mainly two types of clustering:
3. Soft clustering and hard clustering:
4. In soft clustering, one element or one pixel may belong to more than one clusters with different degrees of membership. Hence it is also known as overlapping clustering or soft clustering. Eg: Fuzzy c-means
5. In hard clustering, one element or one pixel only belong to exactly one cluster. Hence it is also known as exclusive or crisp clustering. Eg: K-means
6. The specific criterion to be used for clustering can be intensity i.e on the basis of the same color, and/or it can be based on the same texture and/or it can be based on distance [2].
7. As, In proposed system each pixel of an image will belong to only one cluster. Hence we use hard clustering algorithm i.e K-means algorithm [2].

III. IMPLEMENTATIONS

K-means algorithm:

The steps involved in k-means are:

First of all proposed system will decide number of clusters i.e K. Then proposed system assume the center of these K clusters. Proposed system can take random objects as the initial centroids.

Then algorithm do following three steps till all objects does not change the cluster.

a) Determine the new centroid for each cluster.

b) Determine the distance of each object from newly found centroids.

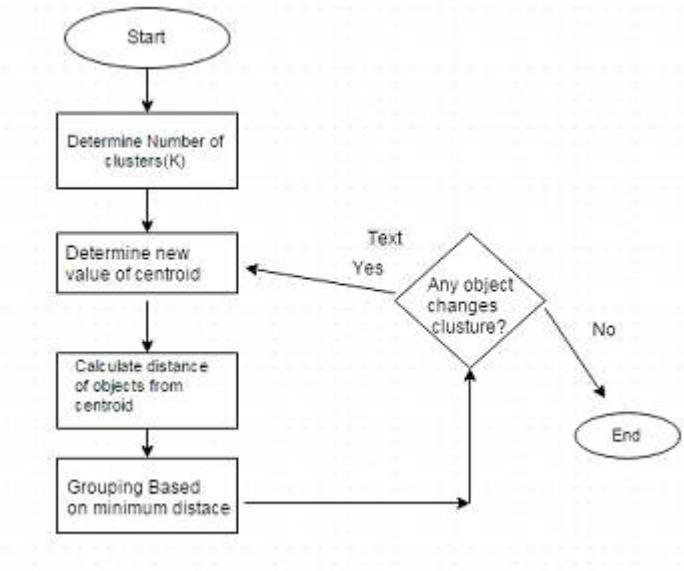

Fig 1: Flow Chart

c) Grouping of each object based on minimum distance. Flowchart of K-means

Distance of object from centroid is calculated by euclidean formula.

K-means on Image:

a) Read the image:

Read the image i.e read the RGB value of each pixel in the image & save it to an array.

b) Determine the numbers of clusters i.e K

For an image it can be takes as No. of colors are there in the image by ignoring variations in brightness.

c) Conversion of image from RGB to L*a*b*:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

Now we have to decide how these colors differ from one another and we have to measure the quantity these visual differences. The L*a*b* can gives us that quantity. This is also known is Lab color space.

One of the main advantage of using L*a*b* model is that it is device independent. L*a*b* space consist of:

L->luminosity layer

a*->indicates the where color falls along the red-green axis.

b*->indicates the where color falls along the blue-yellow axis.

Hence all the color information is present in 'a*' and 'b*'.Difference between two colors can be found out by taking Euclidean distance metric.

d) Apply K-means on the image based on L*a*b* values.

e) Apply till no pixel is changing the cluster.

The sample output after applying K-means Algorithm on one sample image is shown in Fig 2 and Fig 3

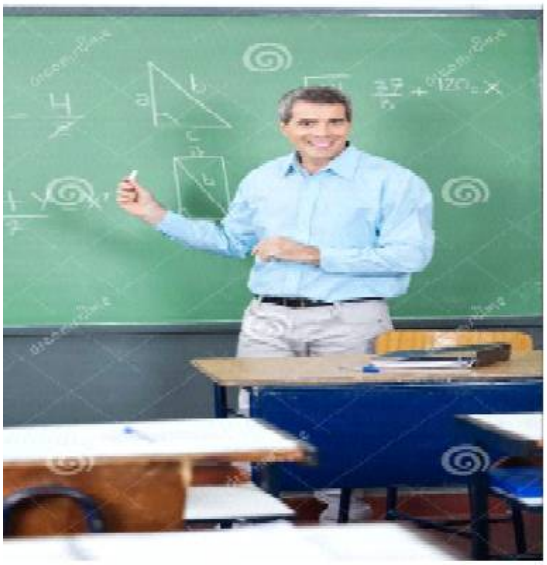

Fig 2: Original Image

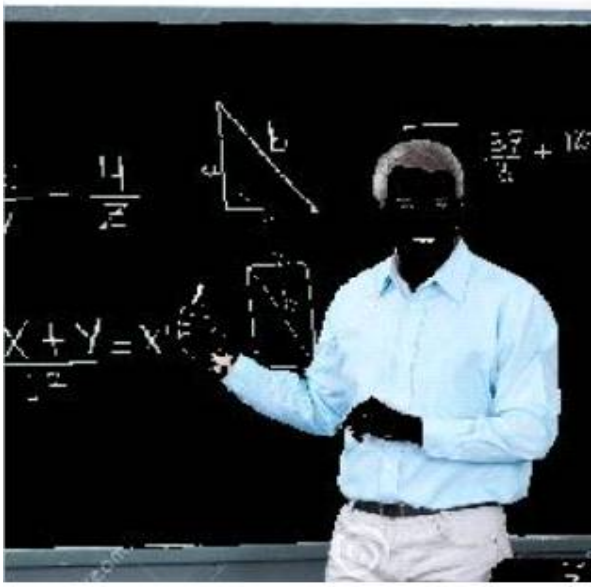

Fig 3: Clustered Image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

WEKA TOOL

The Proposed system will apply K-means clustering algorithm to set of images. Then found clusters of image are stored in the database. Then audio of the video is recovered and converted into text. After collecting all the description video is mined using Weka tool. Weka is a tool consisting of collection of machine learning algorithms for data mining containing any kind of information whether be video or images or audio. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for visualization, data pre-processing, regression, classification, clustering, and association rules. It is also well-suited for developing new machine learning schemes. From extracted text and clustered images System will give appropriate category and provides suitable title for the video.

After applying clustering and mining on set of images from video, system will give appropriate category and provides suitable title for the video.

IV. CONCLUSION AND FUTURE WORK

The Proposed Paper have discussed the system for video mining by content. Video mining by content is a complex subject because it depends of a great number of factors. The proposed system will require a more amount of space complexity as well as time complexity on the server side but will provide integrity and ensure trust of the client regarding video surfing.

REFERENCES

[1].R. V. Patil ,K. C. Jondhale ,"An Efficient k-Means Clustering Algorithm: Analysis and Implementation"July 2002.
[2]Vijay Jumb, Mandar Sohani , Avinash Shrivas ,"Color Image Segmentation Using K-Means Clustering and Otsu's Adaptive Thresholding"February 2014.
[3]Jinghua Wang and Guoyan Zhang,"Video Data Mining based on K-means Algorithm for Surveillance Video" Oct. 2011..

BIOGRAPHY

Dhiraj Paryani, Hariba Dhere, Sumit Lalwani and Nikhil Talreja are students of Computer Engineering Department, V.E.S. Institute of Technology, Chembur, Mumbai. **Mrs. Dimple Bohra** is an assistant professor in the Computer Enginnering Department, V.E.S. Institute of Technology, Chembur, Mumbai. Their research interests are Data Mining, Artificial Intelligence Algorithms,Image Processing etc.