RESEARCH ARTICLE OPEN ACCESS

Discriminative Feature Based Algorithm for Detecting And Classifying Frames In Image Sequences

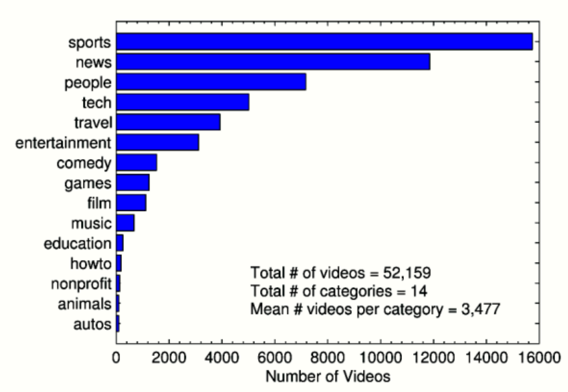
M. Antony Arockia Victoria, R. Sahaya Jeya Sutha
B.E,M.E. Assistant Professor, Department of MCA, Dr.Sivanthi Aditanar College of Engineering,
MCA,M.Phil. Assistant Professor, Department of MCA, Dr. Sivanthi Aditanar College of Engineering

ABSTRACT
This method is used to detect and classify frames in different videos. By detecting frames similar video as input video and related videos are retrieved. Content Based Copy Detection method used to find content related frames from multiple shots. To improve the efficiency of Content Based Copy Detection the videos are cropped that means removing the black bars from horizontal and vertical position. These cropped videos are robust to cam cording and encoded video. Affinity Propagation and exemplar based clustering used to reduce the number of frames in each video. In Exemplar based clustering unique frames are selected from multiple shots and Affinity Propagation used to cluster the unique frames. So it will be useful in detect the frames and compare the input video with all frames. Affinity Propagation uses different similarity metrics to detect the difference between two frames or two videos. Frame classification results are achieved by tiny videos when compared with the tiny images framework. Therefore video frames convert into low dimensional resentation that means resize the frames Frames into 32*32 pixels and concatenating the color channels to reduce the sensitivity in variation. Simple data mining technique that is nearest neighbor method to perform related video retrieval and frame classification. This method can be effectively used for recognizing research that is video as same as to input video will be retrieved and related videos also retrieved.

I. INTRODUCTION

Videos are collected from YouTube's News Sports, People, Travel and technology section. The tiny video database are contains preprocessed vide, the videos are collected from YouTube and fed into some preprocessing steps. first, split video into frames then remove black bars from horizontal and vertical position second step is resize these frames into 32*32 pixels and convert into low dimensional conversion, using this LUV conversion variation sensitivity can be reduced, after that with the help of Affinity Propagation unique looking frames are identified from each video AP Sampling discard all similar frames and retained only unique looking frame. This method can be useful while content based copy detection. Here content based copy detection Used to identify the video as same as input video and related video of input video.

There are main advantages in frame splitting and classification. First, it improves accuracy of related and duplicate frame identification because each and every frame is compared with Input video frame.



Content based copy detection used to detect the same video shots occurring in different. Classification results achieved by tiny videos are compared with tiny image dataset.

Here frames are resized like tiny image so that tiny frames are also compatible with tiny image dataset. This image dataset can be useful in classifying frames into broad category. Same descriptor can be used for tiny video database and tiny image database. To retrieve related video, simple data mining techniques used that are nearest neighbor method it reduces complexity.

Content Based Copy Detection schemes appeared as an alternative to the watermarking approach for persistent identification of images and of video clips.CBCD approach only uses content based comparison between the original video stream and the controlled one. For storage and computational considerations, it generally consists in extracting as few features as possible from the video stream and matching them with the database. Content Based Copy Detection presents two major advantages. First, a video clip which has already been distributed can be recognized. Secondly, content based features are intrinsically more robust than inserted ones because they contain information.

II. EXISTING SYSTEM

To obtain such a large amount of data, they download 80 million images from the web. The images are re-sized to 32*32 pixels before they are added to the dataset. These 80 million tiny images are used to perform object and scene recognition, object localization, image orientation detection, and image colorization. To detect and classify production effect in video sequences a feature based algorithm is used.cuts, fades, dissolves, wipes and caption are detected by this method. The most common production effects are scene break which mark transition from one sequence of consecutive images to another. Cut is an instantaneous transition from one scene to next. A fade is a gradual transition between a scene and a constant image (fade-out) or between a constant image and a scene (fade-in).A dissolve is a gradual transition from one scene to another, in which first scene fades out and the second scene fades in. Another common scene break is wipe in which a line moves across the screen, in this paper we can detect and classifying frames in image sequence. Tiny videos better suited for classifying scenery and sports related activities while tiny image perform better at recognizing object.

A large number of video summarization algorithms have been developed to perform temporal compression [5],[10]. In uniform sampling, frames are extracted at constant interval. The main advantage of this approach is computational efficiency. However, uniform sampling tends to oversample long shots or skip short shots. Another method is intensity of motion sampling, Intensity of motion has also been used as feature vector for describing motion characteristics. Intensity motion is defined as the mean of consecutive frame differences.

$$A(t)=1/xy\sum \left| L(x,y,t+1)-L(x,y,t) \right|$$

Where X and Y are dimensions of the video and L(x,y,t) denotes the luminance value of pixel (x,y) of a frame at a time t.Intensity of Motion key frame selection algorithm is robust to color and affine transformations. These properties make it suitable for Content based copy detection.

III. PROPOSED SYSTEM

Our aim is to obtain a similar dataset from videos, and to explore its applications to video retrieval and recognition. In this paper we proposed a new summarization algorithm that uses exemplar based clustering to select only unique looking key frames. Exemplar based clustering not only captures within visual appearance variations, but also consolidates similarity across multiple shots. In this paper affinity propagation algorithm used to cluster densely sampled frames into visually related groups. Only the exemplar frame within each cluster is retained and rests are discarded. AP sampling is particularly suitable for define what “unique

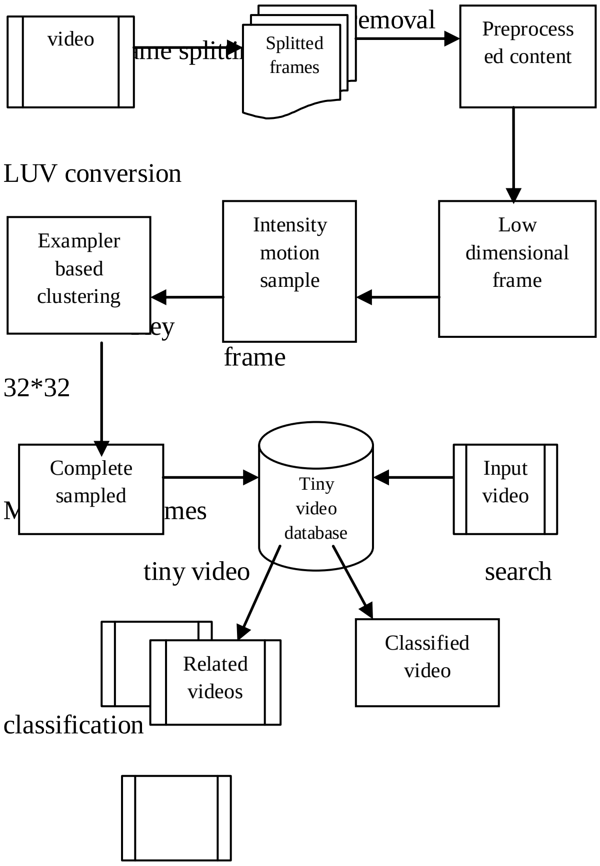


Fig. 2. System Architecture

Looking” means in terms of the same frame similarity metrics that for video retrieval.

Similarities between two frames or two videos are defined by basic distance between two tiny images I_a and I_b as their sum of squared difference;

$$D^2_{ssd}(I_a,I_b)=\sum_{x,y,z} (I_a(x,y,c)-I_b(x,y,c))^2$$

Where I denotes a 32*32*3 dimensional zero mean, normalized tiny video frame or tiny image. We show that recognition performance can be improved by allowing the pixels of the tiny image to shift slightly within a 5-pixels window.

IV. TINY VIDEO REPRESENTATION

A) Video Collection

The Videos were primarily collected in YouTube’s News, Sports, People, Travel and Technology. For each video, we also store all of associated metadata returned by YouTube API. The metadata includes such information as video duration, rating, view count, title, descriptions and assigned label.

B) *Video Preprocessing Procedure*

In this paper videos are preprocessed that is remove black bars from horizontal and vertical position using following formula

$$F(y) = 1/M \sum_{x,c} |I_y(x,y,c)| ,$$

$$Y_{min} = \min[\arg_y \min (f(y)>t)],$$

$$Y_{max} = \max[\arg_y \max (f(y)>t)] .$$

Where I_y is the derivative along y of frame I, which is an M*N*3 matrix. Remove frames that contain more than 80% of pixels of the same color.

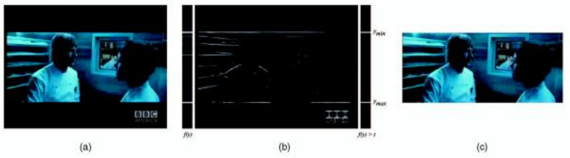


Fig. 2. Removal of black bars

(See Fig 2)The region above y_{min} and below y_{max} is only cropped if they contain at least 80 percent black pixels. Fig. c. shows after horizontal and vertical crop.

C) *Low Dimensional Video Representation*

In this paper Video frames are resized into 32*32 pixels and three color channels are concatenated. This normalized tiny frames are ready to compatible with tiny image frame work, this will improve classification result. A large number of video summarization algorithms have been developed to perform temporal compression [5],[10]. In uniform sampling, frames are extracted at constant interval. The main advantage of this approach is computational efficiency. However, uniform sampling tends to oversample long shots or skip short shots. Another method is intensity of motion sampling, Intensity of motion has also been used as feature vector for describing motion characteristics. Intensity motion is defined as the mean of consecutive frame differences.

$$A(t)=1/xy\sum |L(x,y,t+1)-L(x,y,t)|$$

Where X and Y are dimensions of the video and L(x,y,t) denotes the luminance value of pixel (x,y) of a frame at a time t. Intensity of Motion key frame selection algorithm is robust to color and affine transformations. These properties make it suitable for Content based copy detection.

Fig. 3. Intensity of motion plots with Gaussion filters applied. (a) x= 10 (b) x= 30.

In this paper we proposed a new summarization algorithm that uses exemplar based clustering to select only unique looking key frames. Exemplar based clustering not only captures within visual appearance variations, but also consolidates similarity across multiple shots. In this paper affinity propagation algorithm used to cluster densely sampled frames into visually related groups. Only the exemplar frame within each cluster is retained and rests are discarded. AP sampling is particularly suitable for define what "unique looking" means in terms of the same frame similarity metrics that for video retrieval.

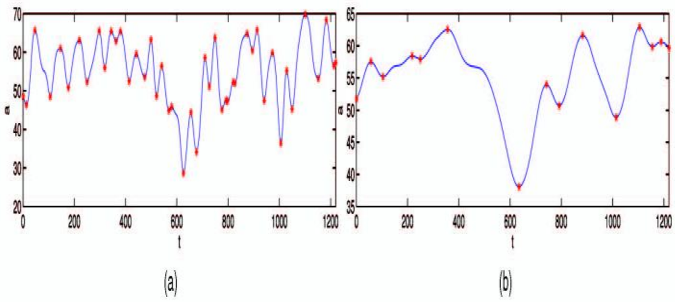
Fig. 4. Comparison of (a) uniform sampling to (b) AP sampling.

The green samples are of a scene that was already sampled once. But the blue samples are not present in uniform sampling. The Red samples are content missing scene from AP sampling.

D) *Frame and Video Similarity Metrics*

Similarity between two frames or two videos are defined by basic distance between two tiny images I_a and I_b as their sum of squared difference;

$$D^2_{ssd}(I_a,I_b)=\sum_{x,y,z} (I_a(x,y,c)-I_b(x,y,c))^2$$



Where I denotes a 32*32*3 dimensional zero mean, normalized tiny video frame or tiny image. We show that recognition performance can be improved by allowing the pixels of the tiny image to shift slightly within a 5-pixels window.

$$D^2_{shift}(I_a,I_b)=\sum_{x,y,c}\min_{|Dx,y|\le w}(I_a(x,y,c) I_b(x+D_{x,y}+D_{y,c}))^2$$

V. CONTENT BASED COPY DETECTION

The aim is to detect the same video shots occurring in different video. our preprocessing steps coupled with the small size of tiny video frame make our descriptors and similarity metrics robust to cam cording, strong recoding, subtitles and mirroring transformation.

A) *Related Video Retrieval Using Tiny Videos*

To find YouTube videos those are related by content (that is sharing at least one duplicate shot) and to evaluate frequency of such occurrences. Precision is defined as the fraction of videos identified as containing duplicate shots. Recall indicates a fraction of related videos found out of all videos with a duplicate shot.

VI. VIDEO CATEGORIZATION

In this paper large data base of videos to classify unlabeled images and video frames into broad categories. We compare our classification results with tiny images.

A) *Labeling noise*

The label for an image could originate from surrounding text which does always describe the image's content. This means that each tiny image is loosely tied to its label. Tags assigned by user with a specific goal of describing the content of videos. A label for video could only apply to a specific segment of videos and completely unrelated to other parts. Therefore many video frames in the tiny video are unrelated to videos label since the user does not indicate which labels apply to which part of the video.

B) *Classification based on WorldNet Voting*

To reduce the labeling noise, we use word net voting scheme. Our goal is to classify "person image", then not only do the neighbors labeled with the person tag vote for the category (E.g. politician, scientist and so on).Videos in our dataset very frequently have more than one label (tag). To ensure that a video with multiple tags gets the same vote as a video's vote evenly across all of its tags. Finally, tiny images and tiny videos can be combined in order to improve precision for some categorization tasks.

C) *Categorization result*

In this paper, we evaluate classification performance for tiny image, tiny videos and both datasets combined. The man made device categorization task includes positive examples of mobile phones, computers, and other technical equipment. For people, we use number of votes for the person noun in the WorldNet tree. The noun "person" has children such as "politician","chemist","reporter","child".while negative examples contain no people in them.

D) *Alternate Video Similarity Metric for Categorization*

Videos with a single very similar frame key frame and multiple completely dissimilar key frames tend to make better neighbors in classifying an input image than videos with multiple moderately similar key frames. This is the case because our AP sampling algorithm picks exemplar key frames and discards all other similar looking frames.

D^2_{shift} of the closest frame in video V to our unlabeled input image I_a, we can determine different distance measure which returns an average distance of the n-closest frames $I_{(1\ldots n)}$ in video V for an input image I_a.

$$D^2_n(I, V) = 1/n \sum_{b=1}^{n} (D^2_{shift}(I_a, I_b))$$

the average distance of n closest frames is simply the distance to the closest frame.

E) *Classification Using YouTube Categories*

Tiny video dataset contains metadata provided by YouTube for each video. Unlike tiny image dataset, which has only one label per image, tiny video stores the video's title, description, rating, view count, a list of related videos and other metadata.

All videos on YouTube are placed into a category. The number of categories on YouTube has increased over the years. Currently videos on YouTube can appear in one of 14 categories. Tiny video dataset contains videos from News Sports, People, Technology, Travel categories.

VII. EXPERIMENTAL RESULTS

For our experimental evaluation,we use Matlab, we can understand the performance of tiny video dataset frame classification and retrieval. Fig. 5 shows false positive and false negative results.

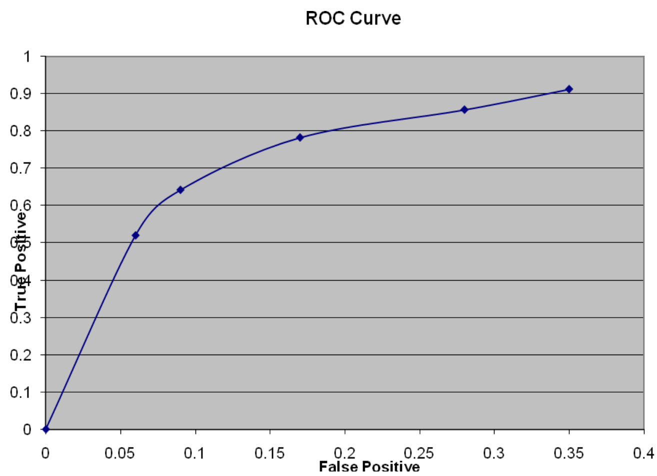


Fig. 5. ROC Curve

Unique frames are clustered using affinity Propagation sampling. Fig. 6. Shows the result of AP sampling.

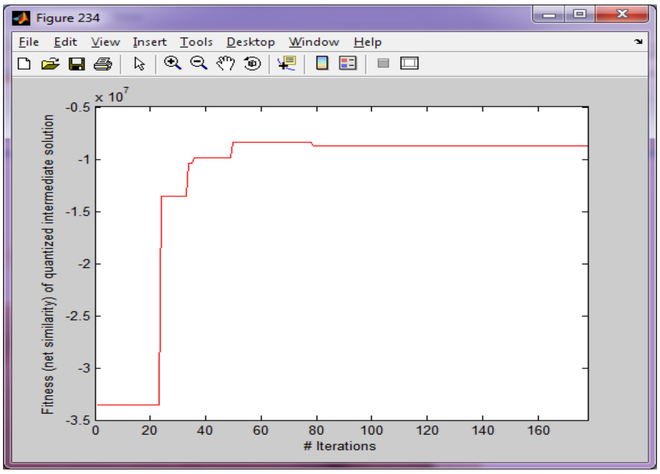


Fig. 6. AP sampling

VIII. CONCLUSION

This paper presents a method for compressing large dataset of videos into compact representation called tiny video. It shows tiny videos can be effectively for content based copy detection. Simple nearest neighbor method used to perform variety of classification tasks. In this paper tiny video dataset designed to compatible with tiny image to improve classification precision and recall. Finally additional metadata in the tiny video database can be used to improve classification precision for some category. The same descriptor used for tiny videos and tiny image. This allows combining two dataset for classification. However RGB color space not used for content based copy detection so in this paper three color channels are concatenated and CBCD method used.

REFERENCES

[1] Aarabi.P and Karpenko.A , (2008) 'Tiny Videos: Non-Parametric Content-Based Video Retrieval and Recognition,' Proc. Tenth IEEE Int'l Symp. Multimedia, pp. 619-624, Dec.

[2] Alexandre Karpenko and Parham Aarabi (2011), 'Tiny Video:A Large Dataset For Nonparametric Video Retrieval and Frame Classification',IEEE Transaction on pattern Analysis and Machine Intelligence,vol 33,No.3.

[3] Buisson.O ,Boujemaa.N , Brunet.V , Chen.L , Gouet , Joly.A , Law-To.J, Laptev.I and Stentiford.F, (2007) 'Video Copy Detection: A Comparative Study,' Proc. Sixth ACM Int'l Conf. Image and Video Retrieval, pp. 371-378.

[4] Buisson,O , Fre´licot.C and Joly.A, (2003) 'Robust Content-Based Video Copy Identification in a Large Reference Database,' Proc. Conf. Image and Video Retrieval, pp. 414-424.

[5] Das.M, Liou.S.P, Toklu.C, (2000) 'Video abstract: A Hybrid Approach to Generate Semantically Meaningful Video Summaries', Proc. IEEE Int'l Conf. Multimedia and Expo, vol. 3, pp. 1333-1336.

[6] Freeman.W.T, Fergus and Torralba,A, (2007) '80 Million Tiny Images: A Large Data Set for Non-Parametric Object and Scene Recognition', Technical Report MIT-CSAIL-TR-2007-024.

[7] Miller's, Mai.K and Zabih.R, (1995) 'A Feature-Based algorithm for Detecting and Classifying Scene Breaks,' Proc. ACM Multimedia Conf., pp. 189-200,.

[8] Miller.J, Mai.K, and Zabih.R, (1999) 'A Feature-Based Algorithm for Detecting and Classifying Production Effects,' Multimedia Systems, vol. 7, no. 2, pp. 119-128.

[9] Niste´r.D and Stewe´nius.H, (2006) 'Scalable Recognition with a Vocabulary Tree,' Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2161-2168.

[10] Shahraray.B,(1995) 'Scene Change Detection and Content-Based Sampling of Video Sequences,' Proc. SPIE Conf., pp. 2-13.

ABOUT THE AUTHORS

M.Antony Arockia Victoria She is a Assistant Professor in Dr. Sivanthi Aditanar College of Engineering, Tiruchendur. She has done her M.E (CSE) in Dr.Pauls Engineering College, Anna University @Chennai in 2012. She received her B.E degree from Raja College of Engineering & Technlogy. Anna University @ Chennai in 2010. She had presented papers in national and International Conferences.

R.Sahaya Jeya Sutha She is presently working as a Assistant Professor of MCA department in Dr. Sivanthi Aditanar College of Engineering, Tiruchendur. She worked as a Academic Counsellor for 5 years in IGNOU, Tuticorin Regional Centre. She has nearly 5 years of experience in fulltime teaching, 6 years part-time teaching and 13 years in Industry. She has done her B.Sc(CS),MCA,M.Phil under Manonmaniam Sundaranar University, Tirunelveli. She developed various software for her Institution while working as a Programmer. She delivered guest lectures in other colleges. She has also organized workshops and conferences. She acted as a coordinator for department association and Computer Education. Her area of interest is Digital Image Processing. She is a life member of Indian Society for Technical education ISTE. She had attended number of Seminars, Workshops, Faculty Development Programme and Conferences.