

Video Annotations and Tagging for Offline Videos using Matroska Container

Prathamesh Parab¹, Prof. Varshapriya Jyotinagar²

P.G. Student, Department of Computer Engineering, VJTI, Mumbai, Maharashtra, India¹

Associate Professor, Department of Computer Engineering, VJTI, Mumbai, Maharashtra, India²

ABSTRACT: Video Tagging is associating information and description with the content of videos like frames, scenes or shots etc. in an efficient manner rather than associating information for the complete video. Video Tagging has found wide applications in many areas and is gaining its importance especially in the fields of social networking, on-line educational lectures etc. with clients having non-uniform processing capabilities. Also with the rise in the multimedia capturing devices the amount of the videos generated are on the rise.

However, most of these videos are monolithic and it is very difficult to identify or directly visit the area in video of immediate interest. Most of the videos are obtained by capturing the live feeds and there are no ways to add more information or rectify any changes necessary about a particular part later. In this paper, we propose a framework for developing an efficient video tagging scheme for the offline videos and storing information in them using Matroska tagging formats which can be user entered (manual tagging) or automated and can be later retrieved at the playback time.

Along with tagging, Video annotations has gain importance in most of the disciplines where the ability to add comments or media-rich notes for specific segments of a video, as in a text editing application, is available in most of on-line video players. However, the same has found limited application in offline videos and are not mostly suitable for research, teaching, learning fields. Also, the annotation systems available are only used for marketing purposes, such as showing advertisements or social commentaries. Here, we suggest a framework for adding, displaying annotation in an offline player and discuss requirements for an annotation system. We propose a standardized pattern for different annotation components, to be added in a video, and store the related information in a video container like Matroska which can be used across different video formats without modifying the actual contents of the video.

KEYWORDS: Video tagging, Video annotation, Matroska, offline video.

I. INTRODUCTION

A video tag is a keyword that you are adding to videos. These keywords describe and highlight the content that is in the videos. Additionally, tags are beneficial for providing additional views about the video. Video Tagging is helpful because video, unlike document, does not include keywords as text. It is the keywords that computers and search engines use to search for content in videos. Computers have a hard time understanding the content of a video and requires a significant feeding of the knowledge about different domains so as to relate to a particular incident or for exploiting visual context in semantic approach for effective analysis of Video. With the rapid increasing of video content on the internet and other multimedia applications, there is an increasing demand in effective technologies for video content analysis.

Today's video sharing and storage services provide an efficient recommendation systems which require incorporating appropriate tags about the contents into the recommendation process. The topic of annotating contents in an efficient, easy, and pervasive way has been an important research area in the field of human computer interaction. Video annotation is a key function to achieve efficient browsing, searching, and recommending videos. While automatic video annotation at the semantic concept level becomes popular topic in the multimedia research community, user-generated video annotation like social tagging is still important to be considered. However, there are some limitations in the tagging approach for the videos. (1) Mostly in current scenarios Video Metadata is limited to information about entire

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

video. Limited efforts are taken to provide information about what exactly are the contents of the video to be included in metadata which can be stored in frame-wise or scene-wise information. (2) Tag storage for the offline videos are still limited to file size, media format, places, no of tracks etc. (Automatically captured video metadata) and there are only few tags available for storing Manually entered video metadata like copyright information, author, keywords, chapters, subtitle information etc.

YouTube [12], is one of the biggest platforms to discover, watch and share video contents for billions of people. It provides YouTube Video Annotation 1 to add additional information to videos in its desktop version. With these mark able components, as we might do in a text editing application, allows a better interactive experience for the viewers. Video Annotations are click-able tools can be added to videos that encourage viewers to take an action, at specific time points in your video. Unfortunately, this ability to add media-rich notes to specific segments of a video is limited in most commonly available online video players. Annotation capabilities in the more popular video players exist primarily for marketing purposes, such as displaying advertisements or for allowing social commentaries. There are very few video players that offers annotation features suitable for teaching, research, learning and other disciplines. Moreover, when the feature is available in a video player, user-generated annotations and corresponding media usually reside in isolated, privately owned information files. Given the success of the annotations in web videos, there is need to standardized integrated temporal annotation solutions for offline video players which will allow users to add and view video annotations in efficient manner.

Most of the videos obtained are from live feed and once we have captured this feed we are not able to modify the obtained video without using any high end professional software. However, with the annotations we will be able to overlay an audio-video segment on top of such unwanted parts and can hide them or notify, acknowledge mistakes with help of overlay objects.

In this paper, we propose a video tagging and annotation attachment framework which can be adopted by offline players for displaying the information. Here we discuss some necessary annotation types and tagging formats. Section III describes the Matroska approach towards tagging of audio-video files. Section IV describes the necessary video annotation and tagging components required for efficient system. Section V shows details of proposed framework. The paper ends with some preliminary conclusions and directions for future work.

II. RELATED WORK

VideoAnnEx [5] is an annotation tool for MPEG video sequences, released by IBM. It parses video sequences and segments them into small shots. Each shot in the video sequence can be annotated with static scene descriptions, key object descriptions, event descriptions, and other lexicon sets. The annotated descriptions are stored as MPEG-7 XML file. Users can add descriptions about each video shot or region in the key frames. VideoAnnEx, however, does not support specification of segments' duration by users. Duration of segments is automatically determined by shot segmentation algorithm. Users need to use its complex ontology editor to enter descriptions. VideoAnnEx runs on Windows machines, and as it stores the information in the MPEG-7 XML file requires the videos to be encoded in the MPEG-7 file.

Similarly other systems like Video Annotation and Analysis system (VANNA) for a Macintosh [6], IBM Efficient Video Annotation (EVA) system, a web based tool for semantic concept annotation of large video and image collections [7], Marquee [8], VAnnotator [9] are all proprietary systems and does not provide any flexibility in terms of the operating system, video formats and also requires the information to be stored on a central file.

The Matroska, an open standard multimedia container [10] format based on Extensible Binary Meta Language (EBML) [11] that envelops an unlimited number of video, audio, picture or subtitle tracks in a single file, has come up with a container format to include tags for the video in the video file itself. It bundles the tagging information in the byte format along with the video file which can be of different video formats. This eliminates the overhead of converting the video in desired encoded format. They also have a standard XML DTD for adding tags to the video but the functionality is limited in terms of providing options only for adding chapters info but not segment related tags.

In our approach, to make the tagging and annotations available offline and to carry this information with the file, we use the tagging available in the Matroska formats for storing tags related to segments of videos having start time and end time. The annotations are stored in XML/Text format and will be attached to the container.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

III. MATROSKA FILES

Exemplar Matroska aims to become the standard of multimedia container formats. It was derived from a project called MCF, but differentiates from it significantly because it is based on EBML (Extensible Binary Meta Language), a binary derivative of XML [10]. EBML enables the Matroska Development Team to gain significant advantages in terms of future format extensibility, without breaking file support in old parsers.

- Tags
 - Tag (about the movie)
 - * SimpleTag
 - TagName = "TITLE"
 - TagString = "Dune"
 - * SimpleTag
 - TagName = "DIRECTOR"
 - TagString = "David Lynch"
 - * SimpleTag
 - TagString = "1984"
 - * SimpleTag
 - TagName = "COMMENT"
 - TagString = "One of the best sci-fi movies"

Fig 1. Example of Matroska Tags Storing the information about the movie

Matroska Tagging:

Here are some real-life examples of tags available with Matroska. For example figure 1. Shows structure for storing the title, director, date of release and comment about a movie on one DVD in XML format.

The segments do not include only the multimedia data but also information (e.g., segment info, tracks, chapters, etc.) needed for constituting the file functional. The general structure of the Matroska file is depicted in the Figure 1. Matroska uses following extensions

- Mkv : Used for Video files, as well those containing audio (movies) or video only
- Mka : Used for audio only files, can contain any supported audio compression format, such as MP2, MP3, Vorbis, AAC, AC3, DTS, or PCM [10].

Matroska is a container format and can contain multiple audio, video tracks simultaneously in a single file allowing attaching the images and the xml like text files. This allows us to choose the appropriate tracks at playback time.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

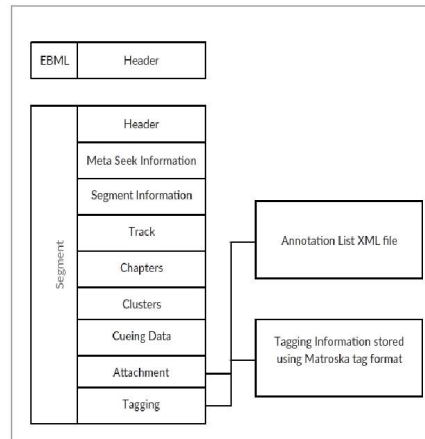


Fig 2. Matroska Files Structure

Figure 2 shows the storage format of the Matroska files along with the metadata information that can be added to the video container. It has a provision to add text files as attachment to the multimedia format which will be used to store the annotation and tagging information. Along with it, the container allows to add multiple video/audio track and even image files as attachments which can be used for annotating the videos.

IV. REQUIREMENTS OF VIDEO TAGGING AND ANNOTATION SYSTEM

With the growth in the media capturing devices, number of video clips and hours of video footage consumed and produced has been steadily growing. Unfortunately the ability to add annotations is limited and not as prevalent nor fully advanced as in the ones available for digital text clients. Very few video players offer annotation features that are suitable for research, teaching or learning. There is a need to add standardized temporal annotation that will allow users to add, retrieve commentaries in a video file. These timeline tags and annotations should not be restricted to simple text only. System should offer the option to cite or cross-reference any media-type as part of the annotations providing links back to original sources through URIs to public resources on web or library information systems. Users should be able to restrict their annotations with predefined peer groups or make them available to public. Likewise, annotators should be entitled to access an entire contextualized archived copy of their contributions along with other publicly available annotations, via a simple file download, in a standardized open file format (txt, xml, html, etc.).

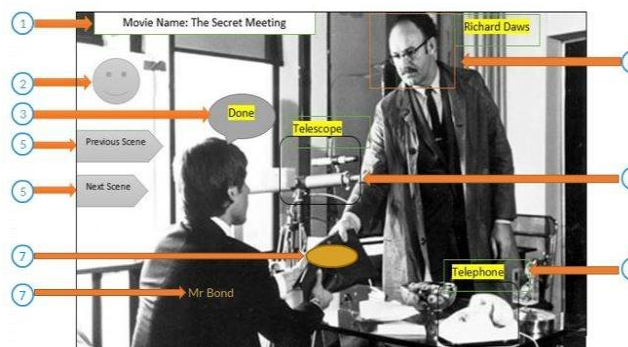


Fig 3. Requirements of a Video Annotation System

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Figure 3 shows the components mapped to the requirements that system must allow users to add as annotations in video which are further illustrated as below.

Interactive components (Click-able components)

1. Annotation or comment of the entire video piece
2. Overlays of images or videos on top of sound visualization frames
3. User-defined custom annotation markers
4. Annotation to highlight a particular area and add captions to field
5. Annotations to seek forward/backward to a particular location within video or to provide links to other videos or to web sites
6. Face tagging to identify the people within the videos

Static Components:

Overlays of user-generated text and graphics associated with an annotation should include:

- Graphics for basic shapes like ellipses, lines, rectangles, polygons and freehand shapes
- User defined stroke and fill properties of graphic objects including colour, stroke weight and transparency
- Short text tags with custom fonts, font weight, size and colour, including all font related attributes.
- Positioning and placement of graphic objects dependent upon the graphed waveform and zooming level

Along with it a time range (beginning and ending time) for each annotation should be provided for exactly marking the starting and ending of the particular annotation.

V. THE PROPOSED METHOD FOR EMBEDDING TAGGING AND ANNOTATION METHOD IN OFFLINE VIDEOS

A high level description of our approach for adding the tagging and annotation information in videos is illustrated in Figure 4.

Initially the raw video file (in any encoding format) is prepared for tagging by finding the existing tagging info if any (for already edited videos). Then for adding each tag or annotation, our system requires the video to be divided into the segments. A segment is part of the video having start time and end time. Users can create a new annotation or tag for the particular segment. Then that segment is marked against the required tag or annotation. The information (parameters for the annotations say text information or the formatting information or background colour/image etc.) about the annotations is stored in the descriptive text format. All the annotation and tagging information is then clubbed into a single XML file format and attached to a Matroska container along with the original raw video. This XML file is parsed at playback time so that appropriate annotations or tags can be regenerated. This reduces overhead required to store the attachment information.

The system allows images, videos and other Medias to be added as overlays which are also attached to the video container using attachments functionality available in the Matroska container. As the tagging and annotation information is stored as text there is a very little overhead in terms of the storage of the tagging and annotation information. Also, the overhead for storing the Matroska headers is low [14].

Users can put annotations into either a specific video frame or a video segment. Annotating a video frame is needed when important objects appear or a specific region of interest is going to start. Because a video frame can be treated as a picture, length of video segments is not considered. Necessary data for annotating a video frame are: start time and words to tag. On the other hand, annotating a video segment is needed when a user wants to specify a region of interest in the video clip. Duration of the video segment becomes important information. Necessary data for annotating a video segment are: start time, words to tag, and length of the video segment. The user interface for our annotation method needs a video player where video playback is displayed and options provided for tagging and annotations. The system

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

has to have a timeline to track the start and end of the segments of the videos indicating the tags and annotations attached with the video track.

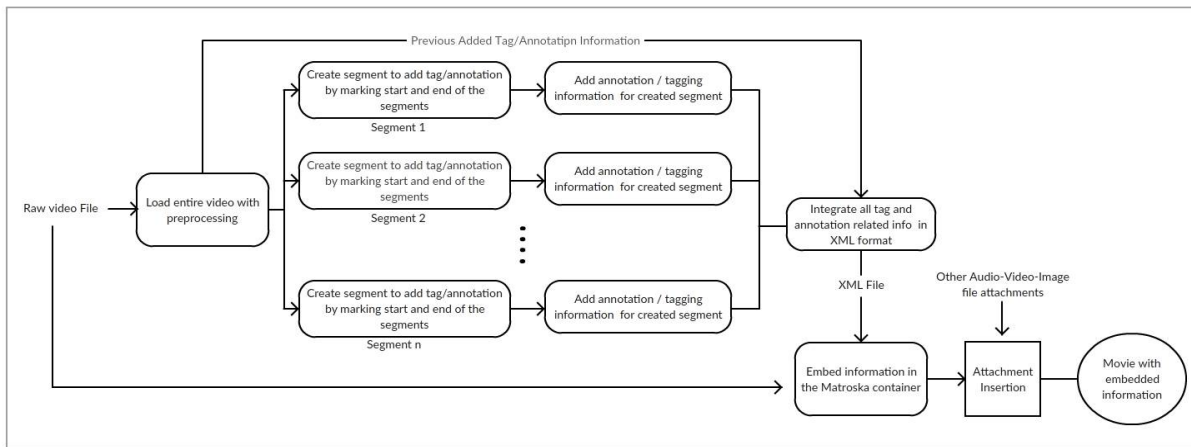


Fig 4. A high level approach of storage annotation and tagging information in the Matroska Container

At the time of playback, the system first parses the annotations XML attached in the video container and must pre-load all the tags and annotations and regenerate the tags as the videos passes by each segment. This parsing information is then rearranged according the starttime and endtime for each tags and annotations to generate a sorted event list. Also, system must preload the tags/annotations in a manner so that they can be loaded as soon as required. As per the event list, for each segments or the frame, the appropriate tag or annotations is loaded and displayed along with the video playback.

The preloading is required due to the seeking functionality associated with most of the video players where we can jump to appropriate sections of the video. As user seeks the video, appropriate annotations must be loaded for that video segment. For extensibility purpose, there must be single structure for all the annotations that should be simply available in xml-like formats.

VI. EXPERIMENTAL RESULTS



FIG 5 (I)

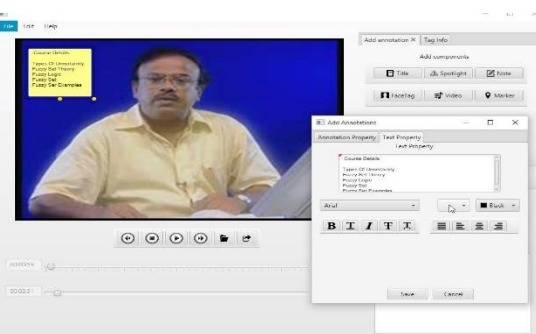


FIG 5 (II)

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

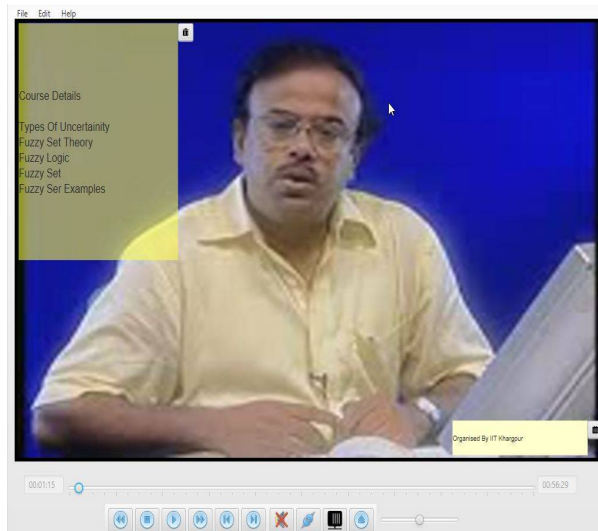


Figure 5 (III)

```
<Video_anno_data>
  <annotation id="1" type="com.vid.comp.Jcomp.Note">
    <starttime>67794</starttime>
    <endtime>3389734</endtime>
    <parameters set="1">
      <StartX>445.05</StartX>
      <StartY>363.05</StartY>
      <Width>137.0</Width>
      <Height>31.0</Height>
      <BgColor>rgba(255,255,204,100.0)</BgColor>
      <DisplayString>Organised By IIT Khargpur</DisplayString>
      <DisplayStringColor>rgb(255,255,255)</DisplayStringColor>
      <Font>Arial</Font>
      <Font_size>10</Font_size>
    </parameters>
  </annotation>
</Video_anno_data>
```

Figure 5 (IV)

To investigate the use of video annotation we developed an annotation system that enables to add the annotation for offline videos. Based upon the approach the system gives an easy to use and learn tool that can be used to add information to the videos. The figure 5(I) shows the video that is to be edited played using a third party VLC media player. Figure 5(II) shows the editor screen shown for adding and displaying the annotations and the final figure 5(III) displays the video generated after the addition of the annotations and displayed using custom video player. The prototype was developed using Java and VLC libraries for video playback hence can be effectively provides portability over different platforms.

The Figure 5(IV) shows the sample XML information used to save the annotations which is attached to the converted output file in MKV format. This xml information is parsed at beginning and the corresponding annotations are generated at runtime over the corresponding time segment.

VII. CONCLUSION AND FUTURE WORK

In this paper, we described a method to add and retrieve the tagging and annotations for videos stored using Matroska file format. We discussed the necessary annotation components and proposed a way to store them as XML files. This method requires a time efficient, interoperable, cross-platform and cross platform implementation for providing additional information about the video contents which will surely help Video sharing services to enhance the quality of searching, browsing, and recommending videos by utilizing these data.

Also, currently our system is aimed on providing manually created tags and annotations. These capabilities can be utilized to automate the process of the tagging. For e.g. identifying a face in the video and automatically tagging it against a person from the database of peoples or tagging the segments from the transcripts, subtitles, text fields in the video [11]. Also the tagging information can be semantically arranged so as to provide more realistic and meaningful searches with the tagging information available. These will help building an efficient and easily available video metadata databases and fasten the process of mark key areas of interest within the videos with minimal efforts.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

REFERENCES

- [1] B. Cunha, O. Machado Neto and M. Pimentel, "MoViA", Proceedings of the 2013 ACM symposium on Document engineering - DocEng '13, 2013.
- [2] B. Cunha, R. Correia and M. Pimentel, "Mobile Video Annotations", Proceedings of the 21st Brazilian Symposium on Multimedia and the Web - WebMedia '15, 2015.
- [3] N. Pitropakis, C. Lambrinouidakis, D. Geneiatakis and D. Gritzalis, "A Practical Steganographic Approach for Matroska Based High Quality Video Files", 2013 27th International Conference on Advanced Information Networking and Applications Workshops, 2013.
- [4] Y. Li, J. Lu, Y. Zhang, R. Li and B. Zhou, "A Novel Video Annotation Framework Based on Video Object", 2009 International Joint Conference on Artificial Intelligence, 2009.
- [5] VideoAnnEx Annotation Tool <http://www.research.ibm.com/VideoAnnEx/>
- [6] B. Harrison, R. Baecker, "Designing video annotation and analysis systems", Proceedings of the conference on Graphics interface '92, p.157-166.
- [7] T. Volkmer, J. Smith, and A. Natsev, "A web-based system for collaborative annotation of large image and video collections: an evaluation and user study", In proc. of the 13th annual ACM international conference on Multimedia (MULTIMEDIA '05), pp.892-901
- [8] K. Weher and A. Poon, "Marquee: a tool for real-time video logging", In proc. of the SIGCHI conference on Human factors in computing: celebrating interdependence (CHI '94), pp. 58-64.
- [9] M. Costa, N. Correia, and N. Guimaraes, "Annotations as multiple perspectives of video content", In proc. of the tenth ACM international conference on Multimedia (MULTIMEDIA '02), pp.283-286.
- [10] Matroska, <http://www.matroska.org/technical/whatis/index.html>
- [11] V. Kamabathula and S. Iyer, "Automated Tagging to Enable Fine-Grained Browsing of Lecture Videos", 2011 IEEE International Conference on Technology for Education, 2011.
- [12] YouTube, <http://www.youtube.com/>
- [13] J. Yi, Y. Peng and J. Xiao, "Exploiting Semantic and Visual Context for Effective Video Annotation", IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1400-1414, 2013.
- [14] Alexander Noé, "Matroska File Format(under construction!)", January 11, 2009
- [15] M. Ames and M. Naaman, "Why we tag: motivations for annotation in mobile and online media", In proc. of the SIGCHI conference on Human factors in computing systems (CHI '07), pp. 971-980.
- [16] E. Baidya and S. Goel, "LectureKhoj: Automatic tagging and semantic segmentation of online lecture videos", 2014 Seventh International Conference on Contemporary Computing (IC3), 2014.
- [17] VLC Media Player, "<http://www.videolan.org/vlc/index.html>"